DISSERTATION

DESIGN AND SYNTHESIS OF HYBRID NANOPHOTONIC-ELECTRIC NETWORK-ON-CHIP ARCHITECTURES

Submitted by

Shirish Bahirat

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2014

Doctoral Committee:

Advisor: Sudeep Pasricha

Wim Bohm
T. W. Chen
H.J. Siegel

ABSTRACT

DESIGN AND SYNTHESIS OF HYBRID NANOPHOTONIC-ELECTRIC NETWORK-ON-CHIP ARCHITECTURES

With increasing application complexity and improvements in CMOS process technology, chip multiprocessors (CMPs) with tens to hundreds of cores on a chip are today becoming a reality. Networks on Chip (NoCs) have emerged as a scalable communication fabric that can support high bandwidth communications in such massively parallel multi-core systems. However, traditional electrical NoC implementations today face significant challenges due to high data transfer latencies, low throughput, and high power dissipation. Silicon nanophotonics on a chip has recently been proposed to overcome limitations of electrical wires. However, designing and optimizing hybrid electro-photonic NoCs requires complex trade-offs and overcoming many design challenges such as thermal tuning, power, and crossing loss overheads.

In this thesis, these challenges are addressed by proposing novel hybrid electro-photonic NoC architectures and novel synthesis hybrid NoC frameworks for emerging CMPs. The proposed hybrid electro-photonic NoC architectures are designed for waveguide-based and free-space-based silicon nanophotonics implementations. These architectures are optimized for low-cost, low-power, and low-area overhead, support dynamic reconfiguration to adapt the changing runtime traffic requirements, and have been adapted for both 2D and 3D CMPs. The proposed synthesis frameworks utilize various optimization algorithms such as evolutionary techniques, linear programming, and custom heuristics to perform rapid design space exploration of hybrid electro-photonic (2D and 3D) NoC architectures and trade-off performance and power

objectives. Experimental results indicate a strong motivation to consider the proposed architectures for future CMPs, with several orders of magnitude reduction in power consumption and improvements in network throughput and access latencies, compared to traditional electrical 2D and 3D NoC architectures. Compared to other previously proposed hybrid electro-photonic NoC architectures, the proposed architectures are also shown to have lower photonic area overhead, power consumption, and energy-delay product, while maintaining competitive throughput and latency. Unlike any prior work to date, our synthesis frameworks allow further tuning and customization of our proposed architectures to meet designer-specific goals. Together, the architectural and synthesis framework contributions bring the promise of silicon nanophotonics in future massively parallel CMPs closer to reality.

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my sincere gratitude to my advisor Dr. Sudeep Pasricha who continuously inspired me and provided vast guidance towards the completion of this research. About six years ago, when I decided to work on my PhD, I visited many professors and universities in Denver area to communicate my interests and how it can line up with the upcoming technological challenges in the next 10 years. I still remember the first meeting I had with Dr. Sudeep Pasricha that was scheduled for about 30 min, however I kept discussing the research possibilities for about 40~45 minutes, after that I never looked back as I was fortunate enough to join his research team. The course and research work I did under the guidance of Dr. Sudeep Pasricha truly helped me to develop expertise from silicon-to-system. His words of wisdom enabled me to learn cutting edge technologies by understanding and addressing many of the challenges faced by industry as well as the research community. By providing focus on critical and valuable research problems and practical solutions, he helped me to get visibility for my research within well-known conferences, journals and institutions. Dr. Sudeep Pasricha is dynamic, enthusiastic, energetic and full of many ideas, which truly makes him one of the greatest advisers and Professors I have known and worked with. The regular brainstorming sessions with Dr. Sudeep Pasricha provided me great insight on many technological areas. I am truly grateful for his excellent guidance and support over the past six years. Without his guidance and help, this research work was definitely not possible. I recall our brainstorming sessions for every research paper or publication that I worked on, and always felt that I was full of ideas to try solving the next set of challenges every time walking out of his office. I performed many “what-if' analysis cases” to obtain the best possible solutions. Dr. Sudeep Pasricha always kept

raising the bar and always pushed me to do better in every aspect of research. I was also fortunate to attend a number of conferences and present my work in front many researchers and the engineering community. The exciting feedback and awards received during these conferences speaks for itself about the high standard that Dr. Sudeep Pasricha adheres to in all aspects of research that includes every major and minor aspect of research. If I look back, not only have I gained technical knowledge but his guidance has also helped to improve my presentation, communication, and writing skills.

I would like to take this opportunity to thank the respected members of my PhD committee, Dr. Wim Bohm, Dr. T. W. Chen and Dr. H.J. Siegel for their insightful feedback and encouragement for my research. The feedback they provided me helped to enhance and broaden my research. I am also thankful to the many students in Dr. Pasricha's MECS lab for their support and help during my PhD: Ishan Thakkar, Yi Xiang, Yong Zou, C Sai Vineel Reddy, Tejasi Pimpalkhute, Nishit Kapadia, Mark Oxley, Dalton Young and Srinivas Desai. Also this list cannot be complete without mentioning Aditi Kulkarni, Jonathan Apodaca and Jay Smith with whom I worked during my course work and publications and received valuable feedback on many aspects about my work. Weekly team meetings managed by Dr. Sudeep Pasricha helped to institutionalize the knowledge through the presentations and discussions with all team members.

Last but not least, I would like to thank my family for their support to continue my research. My wife Pooja, son Ameya and daughter Neha encouraged and supported me to continue my studies and inspired me to keep going while managing multiple priorities with my family, studies, and job. This achievement was not possible without their immense support.

DEDICATION

To my parents, beautiful wife Pooja, amazing son Ameya and wonderful daughter Neha, all of my friends and family

and

to the memory of my grandfather Dr. B. P. Bahirat and father-in-law V.M Kulkarni

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF KEYWORDS

ACK	:	Acknowledge
ACO	:	Ant colony optimization
CFD	:	Computational fluid dynamics
CMOS	:	Complementary metal oxide semiconductor
CMP	:	Chip multiprocessors
CNT	:	Carbon nanotubes
CTG	:	Critical transition graph as an undirected graph CTG
DFS	:	Clock frequency scaling
DVS	:	Dynamic supply voltage
E/O	:	Electrical to optical
FDMA	:	Frequency division multiple access
FSNPI	:	Free space nanophotonic interface
FSOI	:	Free-space on-chip photonic interconnects
GA	:	Genetic algorithm
ITRS	:	International technology roadmap for semiconductors
MPSoC	:	Multi-processor systems-on-chip
MQW	:	Multiple quantum well
MWMR	:	Multiple write multiple read
NACK	:	Not acknowledge
NAS	:	NASA advanced supercomputing
NI	:	Network interfaces

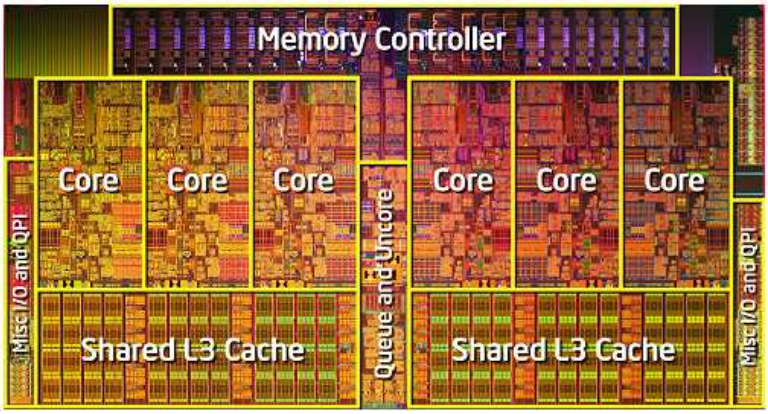
NoC	:	Design of the network-on-chip
O/E	:	Optical to electrical
PARSEC	:	Princeton application repository for shared-memory computers
PCR	:	Photonic concentration region
PRI	:	Photonic region of influence
PSO	:	Particle swarm optimization
QCSE	:	Quantum-confined stark effect
QoS	:	Quality of service
QW	:	Quantum well
RF-I	:	Radio frequency interconnects
SA	:	Simulated annealing
SOC	:	System-on-chip
SOI	:	Silicon on insulator
SWMR	:	Single write multiple read
TDMA	:	Time division multiple access
TIA	:	Trans-impedance amplifier
TO	:	Thermo-optic
TSV	:	Through silicon via
UDSM	:	Ultra-deep submicron effects
UWB	:	Ultra wide band
VCSEL	:	Vertical cavity surface emitting lasers

1 INTRODUCTION

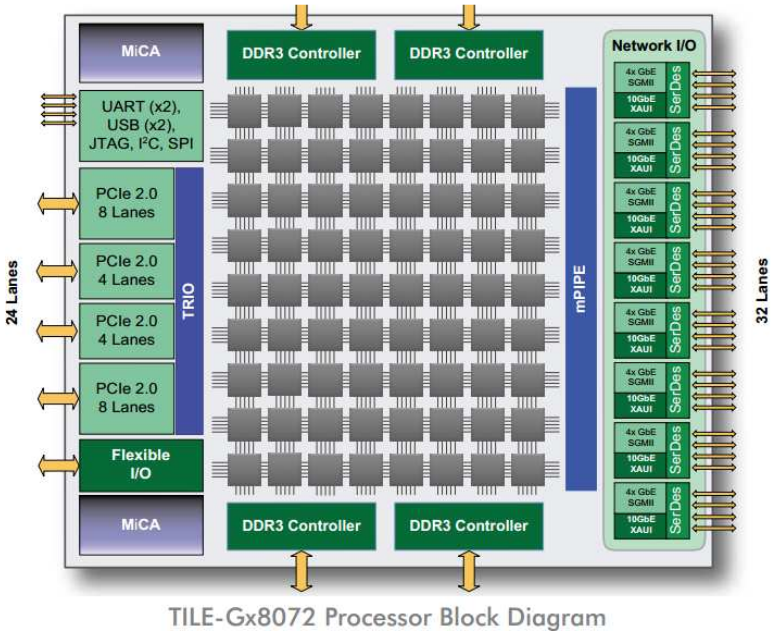
This chapter outlines design challenges for modern chip multiprocessor (CMP) based on ever increasing application performance and power requirements. Network-On-Chip (NoC) communication fabrics in CMPs have led to a paradigm shift from classical bus based architectures to on-chip communication networks; however these network architectures still need to overcome a number of major challenges, which are also discussed in the chapter.

1.1 MOTIVATION

Multicore processors are currently hailed as a universal means to mitigate the power consumption challenge caused by ever increasing frequency scaling to maintain the performance growth required by modern applications. However the real challenge goes far beyond just frequency scaling and encompasses many other complex factors such as technology node scaling and leakage current that consumes power without doing anything useful. Figure 1(a) shows a recent *Intel I7 950 3.06GHz* six core processor. Processors with four to sixteen cores have today become a standard in prevailing computing systems. The trend towards larger multi-core processors is expected to continue, however there are many challenges with scaling the number of cores, such as communication bandwidth, communication power, and communication latency that are becoming limiting factors. Figure 1(b) presents the *Tilera Gx72* processor chip with on electrical mesh NoC that currently consumes around 30~40% pf the total chip power for inter-processor communication [1]. As a result of such high communication overhead, the current fairly explosive growth in the number of cores per processor will have to slow down unless some novel solution addresses many of these challenges in the on-chip communication fabric.



(a)



(b)

Figure 1 Modern processors (a) Intel 6 Core I7 950 3.06GHz, 2012 (b) Tilera Tile-Gx72 system-on-chip, 2013

It is projected that by 2017, as many as 256 cores would be integrated onto a single die. With the advent of such highly parallel chip multiprocessors (CMPs), the design of the interconnection fabric will be crucial to ensure that compute cores are able to communicate with cache banks, memory modules, and other I/O devices with high bandwidths, low latencies and minimal power dissipation. However, electrical communication fabrics today are already

severely constrained due to their long multi-hop latencies and high power dissipation, which will make it practically impossible to stay within the limited on-chip power budget while meeting performance constraints in the near future.

1.2 TECHNOLOGY TRENDS

Continuous CMOS technology scaling over the past several decades has led to transistors becoming faster, smaller in size, and consuming less power. This trend is expected to persist for upcoming years. Moore's law projects *~2×* performance gains through process node technology advancements every *18* to *24* months. The computing world has seen dramatic increases in processor clock frequency from *5 MHz* to *4 GHz* from 1984 onwards, which has so far supported such performance gains. But the increase in frequency is also resulting in increasing power densities to the extent that on chip thermal management has become major a limitation to increasing clock frequency going forward as a means to improve performance.

Figure 2 presents data from Intel® [2] that demonstrates how power density is becoming a critical design issue [1] that needs to be addressed for modern computing systems going forward. This is prompting new way of thinking about how to maintain performance gains without increasing power density. Today, performance gains are being realized through multi-core architectures by doubling the number of cores about every two years. CMPs with multiple cores provide performance gains by exploiting thread level parallelism to complement the traditional instruction level parallelism in uni-core processor systems.

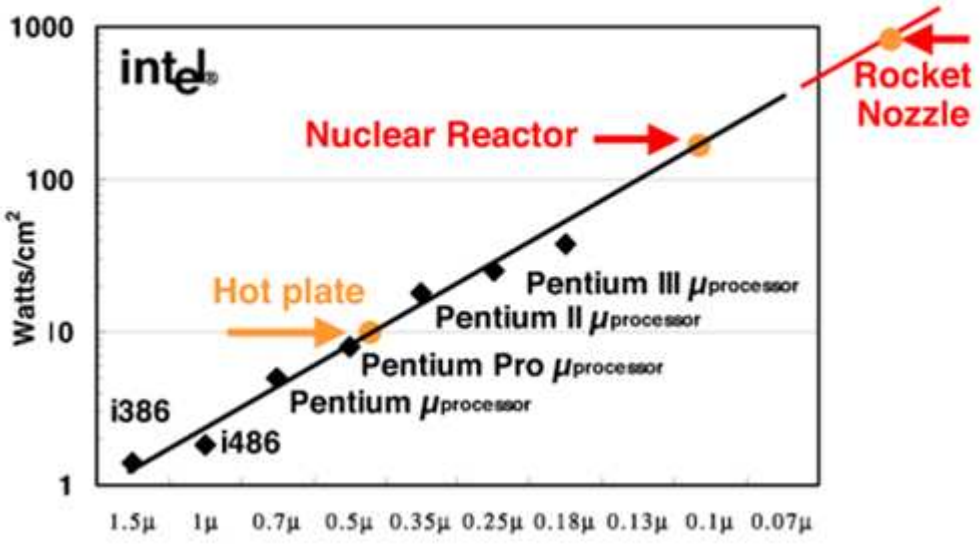
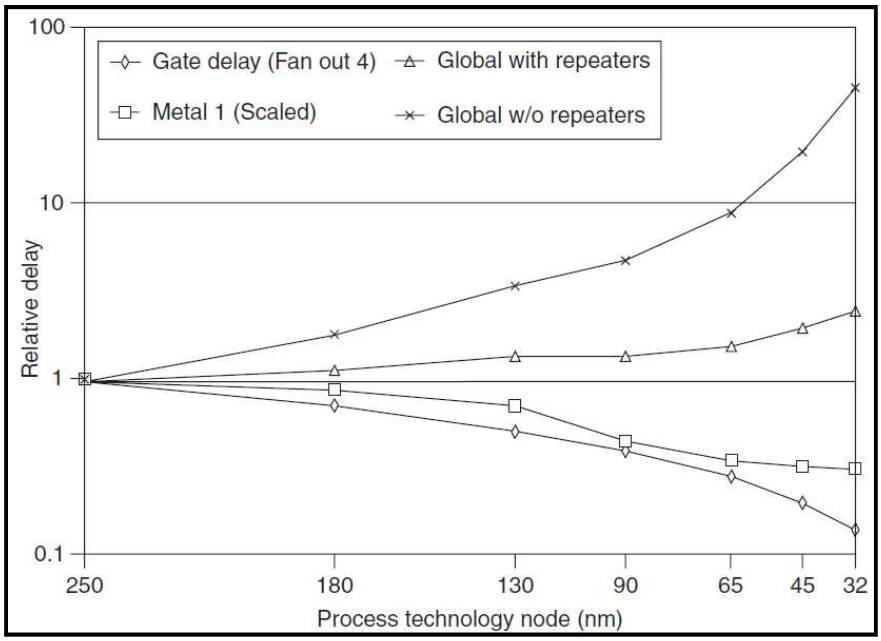
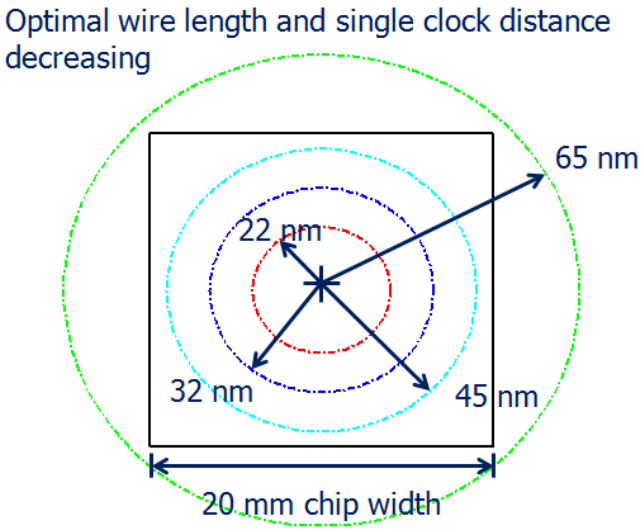


Figure 2 Node size Vs power density data from Intel® [2]

With smaller CMOS technology, gate delay has been decreasing, however the single clock distance travelled for on-chip communication is decreasing. Thus it is taking more clock ticks to communicate from one end of the chip to the other. Also wire delay is characterized by *RC* (Resistance, Capacitance) and this delay has been effectively increasing with technology scaling. As a result gate to wire delay ratio shows an ever increasing gap as shown in Figure 3. Thus it is becoming evident that focus on communication architecture design, customization, and is essential if performance gains are to continue in future generations on CMPs [3].



(a)



(b)

Figure 3 (a) Gate logic delay is decreasing, wire delay is increasing, and gate to wire performance gap is increasing with technology scaling [4] (b) optimal wire length and distance travelled in a single clock cycle is decreasing [2]

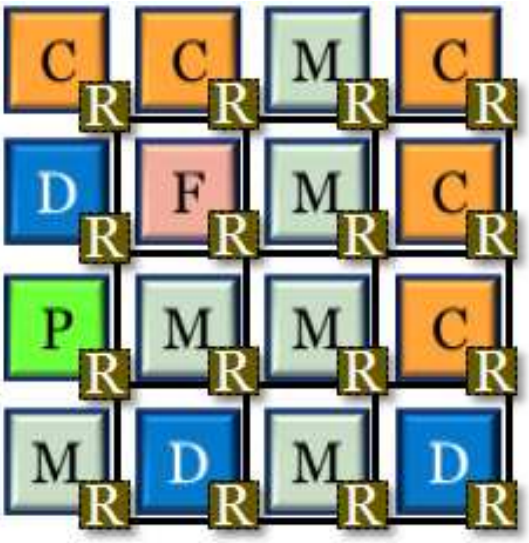


Figure 4 Networks-On-Chip (NoC), C: processor core, M: memory banks D: digital signal processors F: floating point processors P: global power management module

1.3 NETWORKS-ON-CHIP (NOC)

Many current processing systems that include multiple cores encompass a hierarchical or crossbar-type bus-based communication fabric. However bus-based communication architectures do not scale well with the increasing number of on chip cores in terms of bandwidth, clocking frequency and power [5]. Recent years have seen the emergence of a new form of on-chip communication fabric called Network-on-Chip (NoC). A NoC fabric (Figure 4) offers a structured network topology and architecture for on chip communication that reduces the complexity of designing communication fabrics for multi-core systems, providing more scalable latency, power and reliability than bus-based fabrics. NoCs are now therefore being considered as viable options for homogeneous CMPs as well as application-specific heterogeneous multi-processor systems-on-chip (MPSoCs). NoCs have been shown to offer significant benefits in bandwidth, scalability, and reliability compared to traditional hierarchical and crossbar-based shared bus communication architectures in UDSM technologies [6].

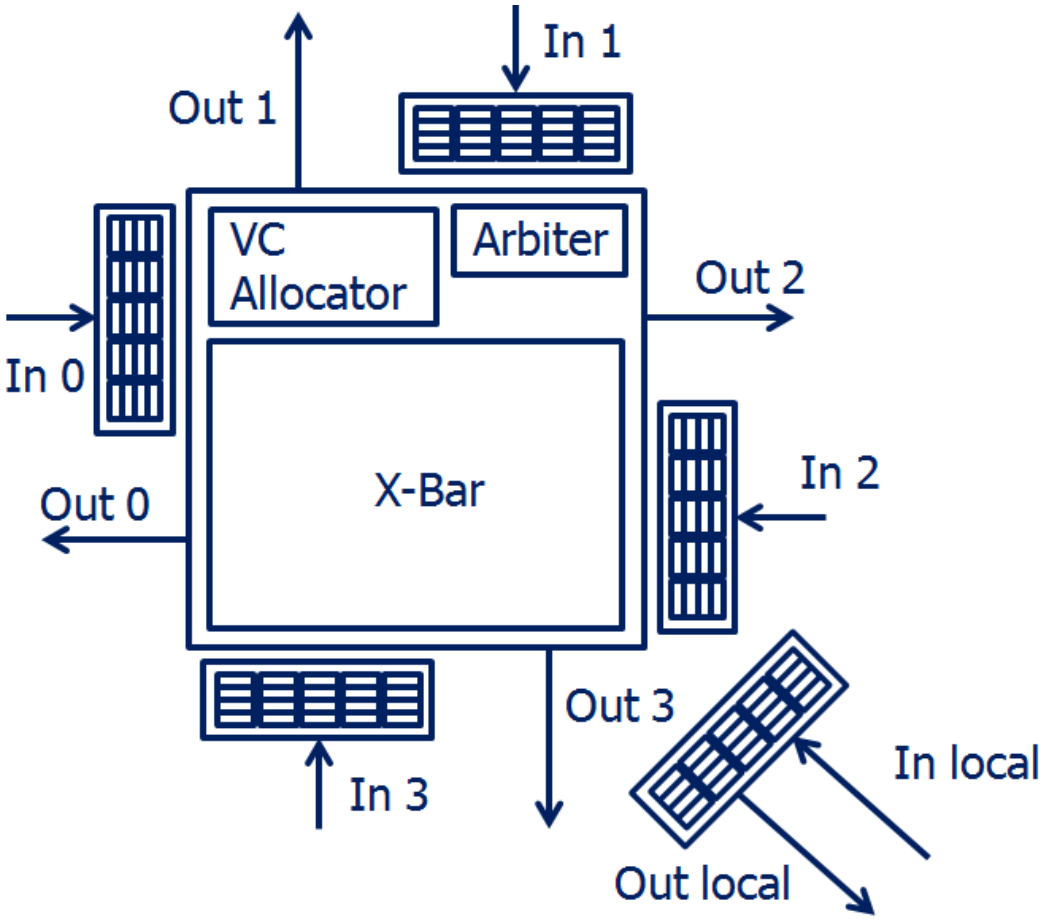


Figure 5 Electrical NoC router

Figure 5 presents an electrical NoC router that includes support for virtual channels (VCs) and wormhole network flow control. This router is designed for two-dimensional regular topologies such as mesh and torus. The NoC router includes 5-input and 5-output ports, where four of the ports are connected to neighboring routers on north, south, east and west directions, and one port is used to connect to a router's local processing core. Every input port supports five buffered FIFO virtual channels that are connected to a crossbar. Routing algorithms such as X-Y routing are enforced by an arbitration unit. The arbitration unit also ensures that there are no

conflicts between each virtual channel and that the arbitration is according to the defined policy. Deadlock control and error handling units can utilize the fifth virtual channel as needed.

In practice, NoC communication fabrics still need to overcome two major challenges that are extremely relevant for emerging CMP applications [7]. *Firstly*, packet switched NoCs must support low latency transfers between cores (especially for real-time applications) but are often unable to meet QoS (Quality of Service) guarantees. Using virtual circuit switching instead of packet switching allows for better QoS management, but unpredictable circuit setup delays still exist. *Secondly*, NoCs must enable low power data transfers. However, the large number of network interfaces, routers, links, and buffers that are part of the NoC fabric lead to a communication infrastructure that consumes a significant portion of the power from the overall CMP power budget. For instance, several prototypes have shown NoCs taking a substantial portion of the system power, e.g., ~30% in the Intel 80-core teraflop chip [1] and ~40% in the MIT RAW chip [8]. Recent studies have suggested that NoC power dissipation is much higher (by a factor of 10×) than what is needed to meet peta- and exa-flop performance levels of future CMPs [7]. This power consumption becomes even worse when the NoC components are designed for adaptive operation with varying runtime application requirements. Power consumption also has a major influence on maximum temperature that determines CMP packaging and cooling costs. Studies [7] indicate that NoC power consumption will continue to ascend as the number of on-chip cores continue to increase. Thus, radical new approaches are required to overcome the power and performance brick walls facing NoCs in the near future [4].

1.4 HYBRID NANOPHOTONIC NOC BASED ON WAVEGUIDES

On-chip photonic communication provides a promising alternative to overcome the abovementioned drawbacks with electrical wires and electrical NoC fabrics. Nanophotonic waveguides have demonstrated bandwidths in the terabits per second range, along with lower access latency and susceptibility to electromagnetic interference [9]. Photonic signaling also has lower power consumption than electrical interconnects for long distance communication, as the power consumption of optically transmitted signals at the chip level is independent of the distance covered by the light [10]. Photonic NoCs were virtually inconceivable with previous generations of photonic technologies. But advances in the field of nanoscale silicon (Si) photonics have enabled the possibility of creating highly integrated photonic CMOS NoC platforms that can send and receive optical signals with superior power efficiencies [11] [12] [13] [14] [15]. In fact, photonic elements have become available as library cells in standard CMOS processes [16]. High-volume capable CMOS photonic transceiver process technology is today being offered by Luxtera [17] [18], in collaboration with ST-Microelectronics [19]. Device simulation libraries [20] now offer simulation of nanophotonic components that can project performance of optoelectronic modulators and waveguide-based silicon photodiodes. Thus, it has now become practical to consider an interconnection network for CMPs built with photonic elements. Such photonic NoCs will likely utilize 3D integration [21] as shown conceptually in Figure 6 with vertical through silicon via (TSV) providing interconnections between the silicon and photonic layers.

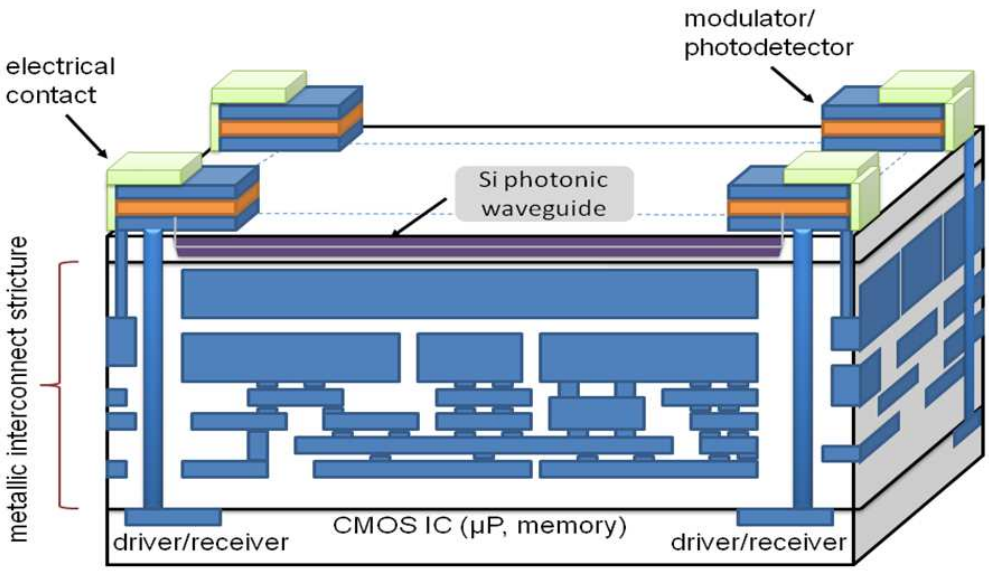


Figure 6 3D IC implementation of a hybrid photonic NoC with cores (bottom layer) and photonic waveguide (top layer) [22]

Recent works have also presented some interesting surveys motivating the need for new technologies on silicon chips. CMP performance scaling up to 20 Giga-flops/Watt will be one of the major challenges going forward [23]. Based on this survey, *(i)* approximately 40% of total power will be used by transistors, *(ii)* 40-50% by the external storage and cooling system thus leaving only *(iii)* 10-20% of the system power for interconnects at all levels. Si photonics is projected to be the leading technology to meet these aggressive requirements. Si photonic technology offers significant and unique advantages in terms of power consumption, access latencies, and high bandwidth for transfers over long distances (>~1mm) on a chip as shown in Figure 7. To realize true benefits of photonic communication, there is a need for intricate balancing of many trade-offs. Figure 7 presents the number of hops that represents distance travelled by a data packet (x-axis) and power consumed (y-axis). Each pink dot represents power consumed by electrical transmission as a function of data packet size, similarly each blue dot represents power consumed by photonic transmission. At smaller distances and small data sizes

electrical communication consumes lower power, but for longer distances photonic communication consumes lower power. Global long distance communication through nano-photonic waveguides however requires two essential enablers: *(i)* buffering and *(ii)* header processing. Both of these are relatively difficult to implement in the photonic realm. Thus a hybrid communication infrastructure with both photonic and electrical signaling becomes necessary.

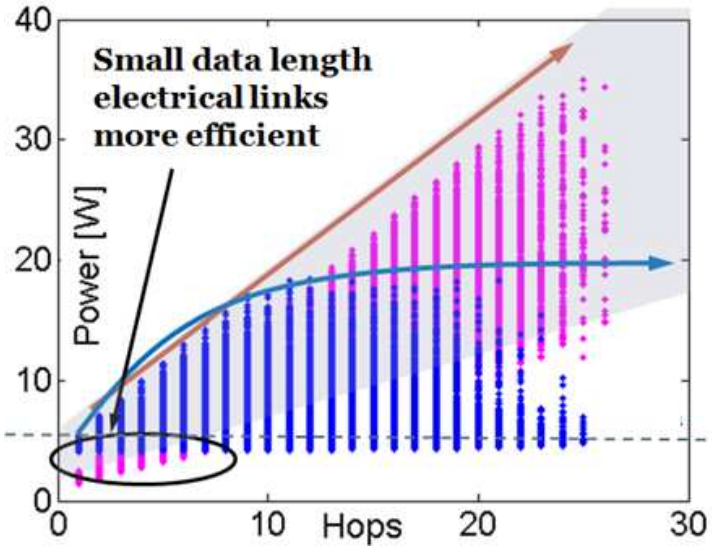


Figure 7 Global/long distance communication, nano-photonic power stays constant/low while electrical communication power increases as a function of distance travelled

1.5 ON-CHIP WAVEGUIDE PHOTONIC COMMUNICATION BUILDING BLOCKS

Figure 8 shows a high level overview of the primary on-chip waveguide photonic transmission components: a multi-wavelength laser light source, resonant modulators/filters, photonic waveguide, and photodetector receivers. Multiple wavelengths of light from a mode-locked, multi-wavelength laser [24] enable wavelength division multiplexing (WDM) that allows several data streams to coexist in the same waveguide, improving transfer bandwidth.

Conventional WDM approaches deploy separate single-frequency lasers with close loop feedback control to stabilize each wavelength that ensures correspondence with the pre-assigned WDM channels [25]. This requires high silicon area and complexity which is critical for an on-chip nanophotonic WDM implementation. An alternative is to use a broadband laser source that supports multiple WDM channels simultaneously where WDM channels are carved out of the broadband spectrum by a passive filter instead of single frequency sources [25]. The main advantage of such an implementation is that the wavelength drift of the source does not influence the system as diffraction grating coupling enables spectrum spread for various wavelengths. Microring resonant modulators [26] convert electrical data signals into light that is propagated through a CMOS-compatible photonic waveguide. The light in the waveguide is eventually coupled into microring filters at the destination that drop the light on photodetectors [27], and thereafter the light signal is converted back into an electrical data signal. Trans-impedance amplifier (TIA) circuits finally amplify analog electrical signals from the photodetector to digital voltage levels. It is also vital for all microring resonators to be thermally tuned (using thermal heater elements) to maintain their resonance under on-die temperature variations.

The topology of the on-chip photonic waveguide has important implications. Photonic waveguides with highly angled structures (such as those commonly found in electrical topologies) may result in significant signal degradation due to bending losses. This degradation is compounded when laying out multiple waveguides for multi-bit parallel transfers on communication links. Consequently, it is preferable to employ simpler topologies, such as a ring topology, that is better suited to the physical characteristics of photonic waveguides. The waveguide is built using a high refractive index silicon on insulator (SOI) material, which has lower pitch and area footprint than low refractive index polymer waveguides such as those used

in [3]. SOI waveguides also have other advantages such as compact modulators that do not require high voltage drive for high frequency operation, the ability to carry light with low losses (on the order of 2–3 dB/cm), and the malleability to be curved with bend radii of $\sim 10\mu m$ [15].

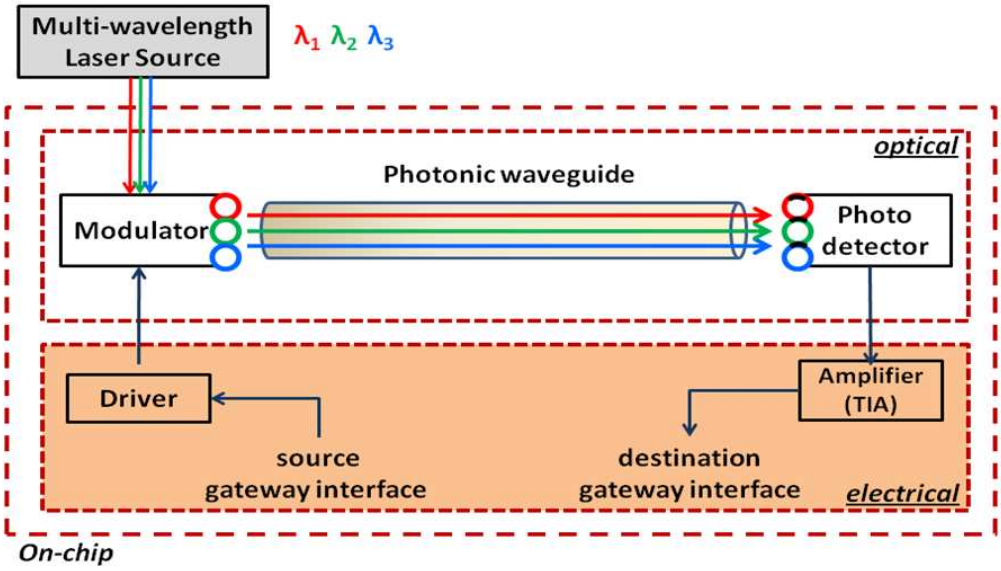


Figure 8 On-chip waveguide photonic transmission components

1.6 ON CHIP FREE SPACE COMMUNICATION BUILDING BLOCKS

Waveguide based nanophotonic communication fabrics that use silicon microring resonators face a few challenges for practical implementation even with recent promising developments. The key challenges for waveguide photonics include: *(i)* high complexity and overhead of thermally tuning microring resonators to ensure proper coupling of wavelengths, *(ii)* high power footprint due to significant waveguide crossing, propagation, and bending losses, *(iii)* need for complex tapered structures and optimized grating couplers with high coupling efficiency, and *(iv) 0.5-3 μm* inter-waveguide spacing requirements to avoid crosstalk that can lead to lower bandwidth density than in optimized electrical wires [28] [29].

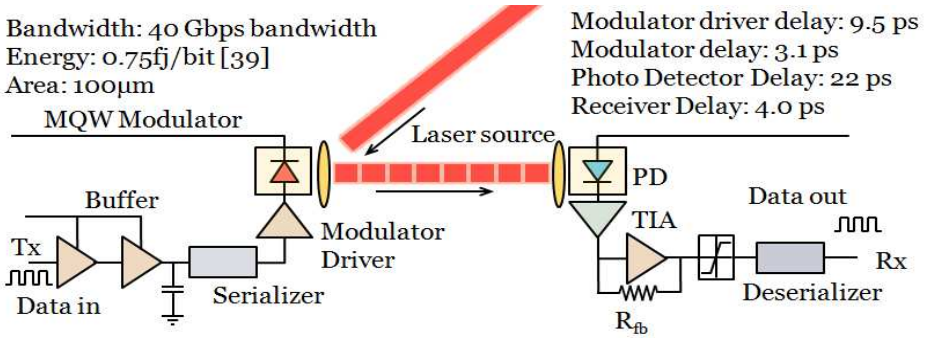


Figure 9 Building blocks of free-space on-chip photonic interconnects

To overcome these challenges with waveguide nanophotonics, free-space nanophotonics based on *GaAs/AlAs* dense Multiple Quantum Well (MQW) devices [30] have recently been proposed as an alternative. In these architectures, light does not travel in a waveguide, instead it is transmitted and reflects off a reflector array (e.g. made of micro-mirrors) to arrive at its destination. Such free-space communictaion can be integrated with standard CMOS fabrication processes and is better suited for high-density optical interconnects due to its small active area and improved misalignment tolerance. MQW devices are projected to consume less than *1* pJ/bit energy and can be configured either as absorption modulators or photo-detectors (PDs). On-chip photonic interconnects utilizing MQWs can operate at *40 Gbps* bandwidth [31] to instantiate single-hop or multi-hop transfers through free-space optical links. MQW modulators provide significant potential to get around the thermal tuning challenges of silicon microring resonators and can be fabricated in various angles to achieve out-of-plane beam steering directions. Figure 9 summarizes the building blocks of on-chip free-space optical interconnects (FSOI) with MQW modulators and PDs. It is also possible to utilize serializer/ deserializer circuits to enable trade-offs between communication power and bandwidth. Figure 10 illustrates a FSOI with logic and photonics planes and vertical through silicon via (TSV) links providing interconnections between

the silicon and photonics layers [30]. 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. The modulated light can be directed through micro-mirrors and micro-lens to transmit data via the free-space medium.

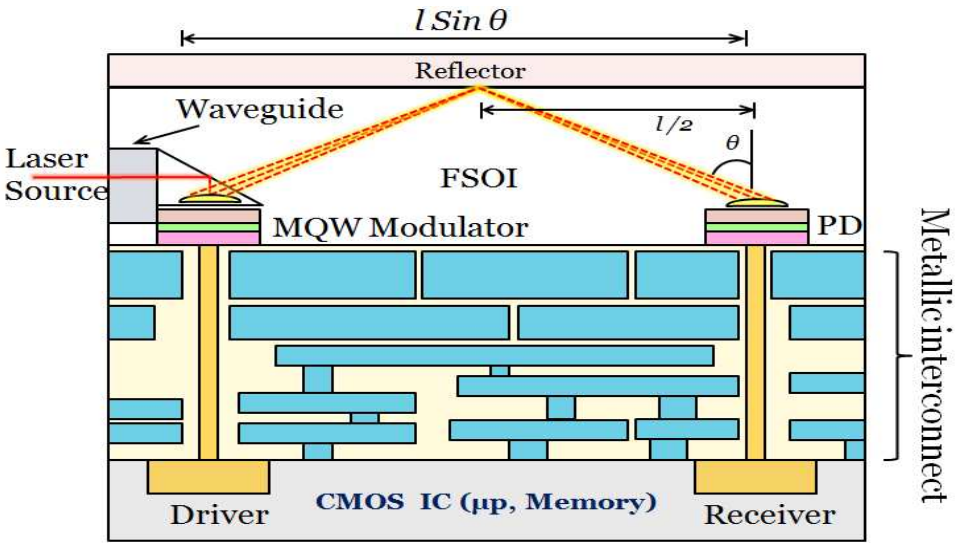


Figure 10 Conceptual view of CMOS integrated free-space on-chip optical link [28]

1.7 3D NOC INTERCONNECTS

While photonic interconnects can reduce on-chip communication bottlenecks, other innovations are required to continually increase core counts on a die, to ensure sustained computation performance increases. Of the several different disruptive technologies that are being investigated for this purpose today, 3D integrated circuits (3D-ICs) with wafer-to-wafer bonding technology is one of the most promising candidates [32] [33] [34]. In wafer-to-wafer bonded 3D-ICs, active devices (processors, memories, peripherals) are placed on multiple active layers and vertical Through Silicon Vias (TSVs) are used to connect cores across the stacked layers. Multiple active layers in 3D ICs can enable increased integration of cores within the same

area footprint as traditional single layer 2D ICs. In addition, long global interconnects between cores can be replaced by shorter inter-layer TSVs, improving performance and reducing on-chip power dissipation. Recent 3D IC test chips from Intel [8], IBM [32], and Tezzaron [33] have confirmed the benefits of 3D IC technology.

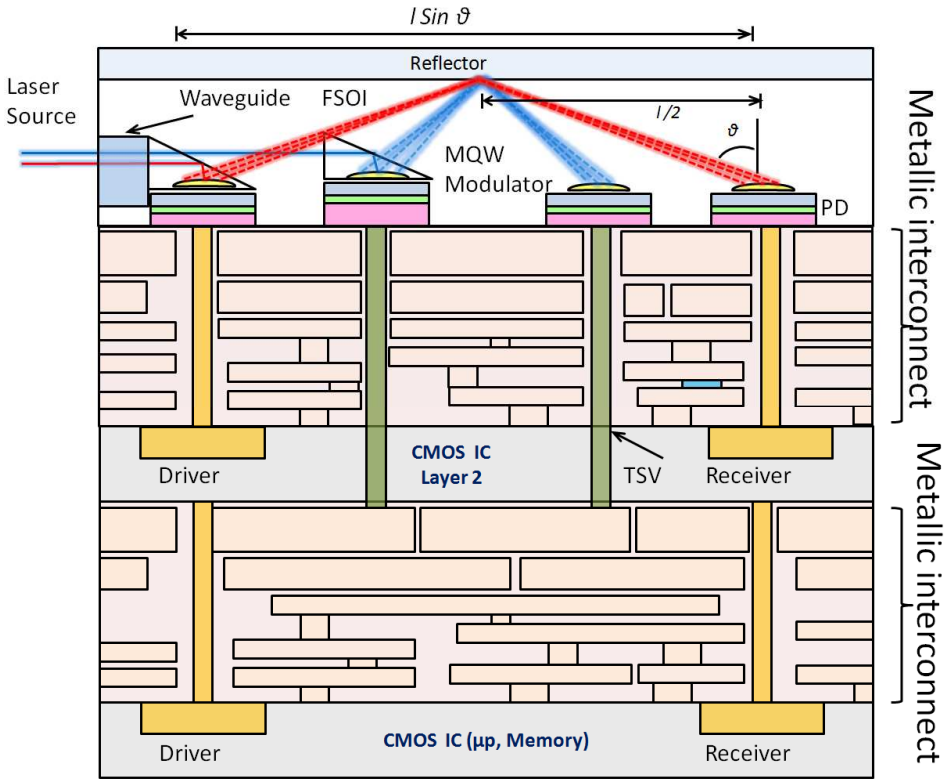


Figure 11 Conceptual view of 3D CMOS IC with logic and FSNPI layers [35]

While 3D ICs are promising, the fundamental power, delay, and noise susceptibility limitations of traditional copper (Cu) interconnects will still limit their achievable improvements. To overcome these limitations, photonic interconnects [29] will be essential components in 3D-ICs to overcome latency and power bottlenecks of Cu interconnects. While several research efforts have individually explored the benefits of photonic interconnects and 3D IC technology,

using 3D ICs as a platform for the realization of hybrid electro-photonic NoCs has not received much attention to date.

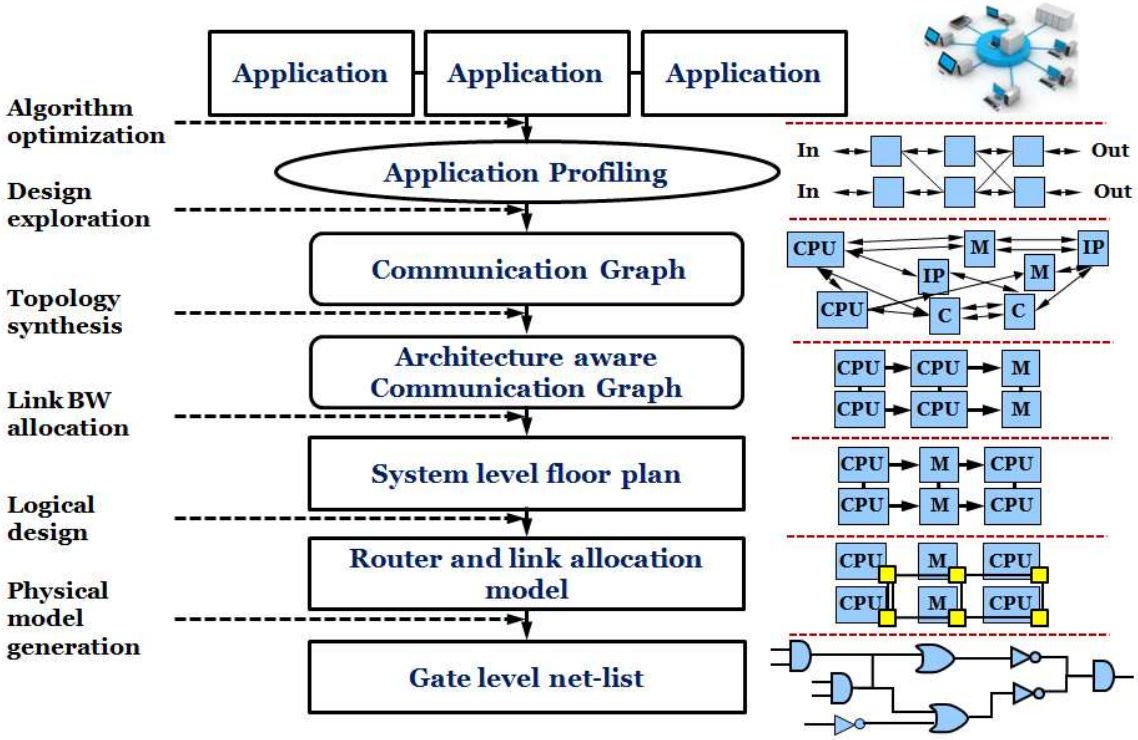


Figure 12 Modern CMP design flow

1.8 HYBRID NANOPHOTONIC-ELECTRIC NOC SYNTHESIS

Figure 12 presents a modern CMP design flow with emphasis given to on-chip communication optimization. The inclusion of photonics in this flow will require many changes. Significant recent research [36] [37] [38] [39] [40] [41] [42] [43] [44] has focused on developing hybrid photonic NoC architectures that optimize local and global communication distribution between electrical and photonic links. However, automated optimization of hybrid photonic NoCs for application-specific performance is an NP-hard problem and has yet to be addressed. Synthesis techniques used for application-specific electrical NoC fabrics [45] [46] [47] [48] [49]

[50] [51] [52] are not directly applicable to hybrid nanophotonic-electric fabrics that possess a more complex search space. Hybrid nanophotonic-electric fabrics require determining values for a diverse set of unique parameters such as wavelength division multiplexing (WDM) density, number of photonic uplinks, serialization degree, etc. in order to maximize communication performance-per-watt [53]. A photonic interconnect with $n=256$ waveguides and $m=256$ wavelengths will require exploring $n+(n)^2+(n)^3+\ldots+(n)^m$ configurations to find the most power efficient solution that also meets performance goals. This is most certainly practically exorbitant [54]. Moreover, these two parameters are just a small subset of the much larger set of parameters that must be explored during hybrid NoC optimization. Finding the best solution for such a combinatorial optimization problem that is known to be NP-hard could take years if searching through the entire solution space, even with leading-edge supercomputing technology today. Thus, application-driven synthesis of hybrid nanophotonic-electric NoCs will become increasingly important as on-chip core counts increase, but this problem has not yet been addressed in prior work by researchers. There is a need to design polynomial-time heuristics that permit us to identify and search through a relevant portion of the solution space in a tractable amount of time to find a near optimal solution [53] [55]. Greedy heuristics are unlikely to find good quality solutions due to their inclination for getting stuck in local optima. In contrast, non-greedy search heuristics such as Simulated Annealing (SA) [56] operate through repeated transformations and have the hill-climbing ability to escape local optima by allowing acceptance of worse solutions within the evaluation process. A population of solutions being simultaneously manipulated is one of the major differences between the SA and traditional greedy search algorithms. Approaches based on SA and other non-greedy iterative algorithms have proven

highly effective in recent years for several hard problems in the realm of VLSI physical design, such as partitioning and placement [57].

1.9 CONTRIBUTIONS

In this thesis, we present novel hybrid nanophotonic-electric network-on-chip (NoC) architectures and synthesis techniques to optimize hybrid nanophotonic-electric NoC architectures for high performance and low power consumption. The main contributions of our work can be categorized in two key areas: architectural and algorithmic techniques.

Our first contribution is *METEOR*, a novel hybrid nanophotonic-electric NoC architecture that utilizes a low overhead photonic ring waveguide to complement a traditional 2D electrical mesh NoC. *METEOR* includes many novel features such as photonic region of influence (PRI), single write multiple read fast reservation channel and multiple write multiple read (SWMR-MWMR) low power data channels, and serialization for gateway interfaces. Experimental results indicate a strong motivation for considering the *METEOR* hybrid nanophotonic-electric NoCs for future CMPs, demonstrating as much as a 13× reduction in power consumption as well as improved throughput and access latencies, compared to traditional electrical 2D mesh and torus NoC architectures.

An enhancement to the above architecture is proposed to support multiple use-case chip multiprocessor (CMP) applications that require adaptive on-chip communication fabrics to cope with changing use-case performance needs. The proposed *UC-PHOTON* architecture is a novel hybrid nanophotonic-electric NoC communication architecture optimized to cope with the variable bandwidth and latency constraints of multiple use-case applications implemented on CMPs. Detailed experimental results indicate that *UC-PHOTON* can effectively adapt to meet diverse use-case traffic requirements and optimize energy-delay product and power dissipation,

with scaling CMP core counts and multiple use-case complexity. For the five multiple use-case applications we explored, *UC-PHOTON* shows up to 46× reduction in power dissipation and up to 170× reduction in energy-delay product compared to traditional electrical NoC fabrics, highlighting the benefits of using the novel communication fabric.

We extended the above work to address scalability issues with 2D ICs and demonstrated a novel multi-layer hybrid nanophotonic-electric NoC fabric called *OPAL* for 3D ICs. Our proposed hybrid nanophotonic-electric 3D NoC architecture combines low cost photonic rings on multiple photonic layers with a 3D mesh NoC in active layers to significantly reduce on-chip communication power dissipation and packet latency. *OPAL* also supports dynamic reconfiguration to adapt to changing runtime traffic requirements, and uncover further opportunities for reduction in power dissipation. Experimental results and comparisons with traditional 2D NoCs, 3D NoCs, and previously proposed hybrid nanophotonic-electric NoCs indicates a strong motivation for considering *OPAL* for future 3D ICs as it can provide orders of magnitude reduction in power dissipation and packet latencies.

As waveguide based hybrid nanophotonic NoCs are constrained due to high thermal tune up power, waveguide crossing losses, we proposed a heterogeneous free space photonics based hybrid nanophotonic-electric NoC architecture with multiple quantum well (MQW) devices and flip chip bonding. The photonic links, using micro-mirrors through free-space can be used to achieve single-hop direct communication links. The proposed architecture includes innovative mechanisms to address free-space collisions and combines single and multi-hop transfers to trade-off performance with power dissipation. We extended this architecture to 3D ICs and the resulting architecture allows for scalable and energy-efficient transfers in 3D ICs.

Any NoC architecture requires optimizations (synthesis) to allow tuning to application-specific characteristics. To date, prior work on automated NoC synthesis has mainly focused on electrical NoCs. For the first time, we proposed a suite of techniques for effectively synthesizing hybrid nanophotonic-electric NoCs with regular topologies by formulating and solving the synthesis problem using four search heuristics: *(i)* Ant Colony Optimization (ACO), *(ii)* Particle Swarm Optimization (PSO), *(iii)* Genetic Algorithm (GA), and *(iv)* Simulated Annealing (SA). Our experimental results reveal significant promise for the ACO and PSO based heuristics, with PSO achieving an average of *64%* energy-delay improvements over GA and *53%* over SA; and ACO achieving *107%* improvements over GA and *62%* over SA.

Finally, we have designed synthesis frameworks that can synthesize hybrid nanophotonic-electric NoCs with irregular topologies based on free-space photonics. The *HELIX* framework synthesizes application-specific hybrid nanophotonic-electric NoC architectures. This framework is also extended to 3D ICs. The resulting *3D-HELIX* framework synthesizes application-specific hybrid nanophotonic-electric 3D NoC architectures. These frameworks are the first to attempt to optimize hybrid nanophotonic-electric NoC architectures based on free-space photonics.

1.10 OUTLINE

The research presented in this thesis is organized as follows. Chapter 2 discusses recent literature survey in the domain of 2D and 3D traditional electrical and nanophotonic NoC architectures as well as the prevailing NoC synthesis efforts that have been mainly focused on electrical NoCs. Chapter 3 describes our novel hybrid nanophotonic-electric ring-mesh NoC (*METEOR*) that utilizes a configurable photonic ring waveguide coupled to a traditional 2D electrical mesh NoC and analyzes performance based on parameter variations. Chapter 4

describes *UC-PHOTON*, a hybrid nanophotonic-electric NoC architecture optimized to cope with the variable bandwidth and latency constraints of multiple use-case applications implemented on CMPs. Chapter 5 presents a novel multi-layer hybrid nanophotonic-electric NoC fabric called *OPAL* for 3D ICs that combines low cost photonic rings on multiple photonic layers with a 3D mesh NoC in active layers to significantly reduce on-chip communication power dissipation and packet latency. Chapter 6 describes a suite of techniques for effectively synthesizing hybrid nanophotonic-electric NoCs with regular topologies. The synthesis problem is formulated and solved using several design space search heuristics. Chapter 7 describes the *HELIX* framework for application-specific synthesis of hybrid nanophotonic-electric NoC architectures with irregular topologies that combine electrical NoCs with free-space photonic links. Chapter 8 extends the *HELIX* framework to application-specific synthesis of hybrid nanophotonic-electric NoC architectures with irregular topologies in 3D ICs. Chapter 9 summarizes the conclusion of this thesis and lists some directions for future work.

2 LITERATURE SURVEY

This chapter describes key published work in the area of NoC architectures and synthesis techniques. We have made an effort to evaluate pros and cons for each approach presented in these works and also have discussed how the work presented in this thesis is different or complements previously published architectures and synthesis algorithms with an emphasis on the critical problems we are trying to address.

2.1 2D AND 3D ELECTRICAL NOC ARCHITECTURES

To overcome the power and latency limitations of traditional electrical NoC fabrics, several research efforts in recent years have focused on architectural customization of NoCs. These efforts aim to optimize performance and power constraints. For instance, [58] presents an interesting work to reduce communication latency and power dissipation in NoCs by inserting long links between certain cores. Express virtual channels are proposed in [59] that allow packets to bypass intermediate routers along their path, to reduce packet latency. Hybrid circuit-packet switched NoCs have been explored in [60] and shown to enable viable trade-offs between power and performance. Several works have also explored circuit-level techniques to reduce interconnect power, such as low swing signaling, power-optimal repeater design and insertion, current mode signaling, and pulsed transmission [61] [62] [63] [64] [65] [66]. Despite these and other architectural and circuit level advances in the field of NoC design, the International Technology Roadmap for Semiconductors (ITRS) [4] projects that in the near future, *the fundamental limitations of copper-based electrical interconnects will become a serious bottleneck*, requiring the exploration of innovative new technologies to sustain the evolution of

VLSI technology. On-chip networks in the Tilera 72 core [8] and Intel 80 [1] chips consume about 30-40% of the total power. This trend is expected to continue as more wire density becomes available within future process technologies [4]. Thus the expected on-chip network power will continue to rise as we scale to hundreds of cores on a single IC.

Several works in recent years have as a result begun exploring novel on-chip interconnect technologies. Such technologies include carbon nanotubes (CNTs) [61] [67] [68] [69] [70] and multi-band RF transmission lines and wireless interconnects (RF-Is) with CMOS ultra wide band (UWB) technology wireless links [71] [72]. However, CNT fabrication is not yet mature and has serious practical concerns to overcome. RF-Is and wireless interconnects require high operating frequencies in the range of hundreds of GHz to THz. Complex RF-I frequency division multiple access (FDMA), transmission lines, and on-chip antennas requires high area, power, verification, and implementation costs. Nonetheless, these technologies are quite promising, and may become more viable in the near future.

Over the last several years, there has also been a growing interest in 3D ICs as a means to alleviate the interconnect bottleneck currently facing 2D ICs. Three dimensional NoCs (3D NoCs) [33] [60] [73] [74] extend traditional 2D NoC fabrics across multiple active layers. Such 3D NoC fabrics are attractive as they can allow long global links to be replaced by much shorter inter-core through silicon vias (TSVs) that can enable low latency and high bandwidth data transfers between cores. A key challenge with 3D ICs is their high thermal density due to multiple cores being stacked together, that can adversely impact chip performance and reliability. Therefore several researchers have proposed thermal-aware floorplanning techniques for 3D ICs [75] [76] [77]. A few researchers have explored interconnect architectures for 3D ICs such as 3D mesh and stacked mesh NoC topologies [74] and a hybrid bus-NoC topology [73]. Some recent

work has looked at decomposing cores (processors [78], NoC routers [79], and on-chip cache [80]) into the third dimension which allows reducing wire latency at the intra-core level, as opposed to the inter-core level. Circuit level models for TSVs were presented in [81].

2.2 PHOTONIC NOC ARCHITECTURES

The concept of photonic interconnects for on-chip communication was first introduced by Goodman et al. [82]. Several works in recent years have explored inter-chip photonic interconnects [36] [83] [84] [85] [86] [87] [88]. With advances in the fabrication and integration of photonic elements on a CMOS chip in recent years, several works have presented a comparison of the physical and circuit-level properties of on-chip electrical (copper-based) and photonic interconnects. [89] [29] [90] [91] [92] [93] [94]. In particular, [90] compared simple photonic and point-to-point links using a spice-like simulator. In [91] photonic and electrical clock distribution networks were studied using physical simulations, synthesis techniques, and predictive transistor models. Both works studied power consumption and bandwidth, and highlighted the benefits of on-chip photonic interconnect technology. Intel's Technology and Manufacturing Group also performed a preliminary study evaluating the benefits of photonic intra-chip interconnects [92]. They concluded that while photonic clock distribution networks are not especially beneficial, wavelength division multiplexing (WDM) based on-chip photonic interconnects offer clear advantages for intra-chip communication over copper in UDSM process technologies. Device level work by [95] introduced novel utilization of CMOS-compatible silicon photonic materials. Polycrystalline silicon and silicon nitride combination demonstrates advantages over traditional crystalline silicon platform. Some of the key benefits include lower waveguide propagation loss, mitigation of waveguide crossing and insertion loss advancing

development of photonic network topologies with unforeseen performance capabilities. Other work from industry and academia has been focusing on photonic device fabrication, for components such as gigascale modulators [96], photodetectors [97], switches [98], couplers, buffers, on-chip waveguides and on-chip wave division multiplexing (WDM) devices [99].

In addition to comparisons between photonic and electrical interconnects at the physical and circuit levels, and device level innovations, a few recent works have explored the system level impact of using photonic interconnects and proposed hybrid nanophotonic-electric NoC topologies [3] [36] [37] [38] [40] [86] [100] [101] [102] [103] [104] [105]. In [101], photonic ring interconnects are considered to improve the latency response and reduce power consumption over electrical bus-based communication architectures. The architecture utilizes electrical bus based flow control that increases communication latency and power dissipation. Hybrid photonic torus architectures [38] [102] are generally comprised of a photonic torus connected to a topologically identical electronic control network that controls its operations and executes exchange of short messages. However, these architectures with photonic torus topologies have significant waveguide crossing losses and photonic layer area complexity. Also, electrical packet switched network based path setup and teardown leads to increased transfer latency and power dissipation in these architectures. The Phastlane packet switched mesh network [103] encompasses a low latency optical crossbar with simple pre-decoded source routing to transmit cache-line-sized packets over several hops in a single clock cycle under contention less conditions. In [106] the PROPEL architecture is proposed that strikes a balance between the electrical and optical realms, by implementing nanophotonic interconnects for long distance communication and electrical switching for routing and flow control. PROPEL implements a nanophotonic crossbar with an electrical mesh topology where the number of distinct

wavelengths required is equivalent to the maximum number of tiles for the architecture. In [40], the Corona all-optical crossbar topology is proposed, with photonic waveguides configured in a token based Multiple Writer Single Reader (MWSR) configuration. The architecture has high photonic layer complexity (e.g., more than a million resonators required for implementation), lacks path diversity and makes use of expensive electro-photonic and photo-electronic conversions even for local transfers, which is inefficient. Similar to [40], in [107] an all-optical network is proposed, 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. In [37] the Firefly topology is proposed which uses a hierarchical crossbar NoC topology with clusters of nodes connected through local electrical networks, and nanophotonic links overlaid for global, inter-cluster communication. The photonic waveguides in the architecture are configured in Reservation-assisted Single Writer Multiple Reader (R-SWMR) configuration. However the architecture has high implementation overhead and no support for controlling the distribution of traffic among the electrical and photonic paths. In [105] bufferless photonic Clos networks are proposed with a novel scheduling algorithm to solve the Clos network routing problem. In [108] an all-optical control architecture is proposed with a minimal deterministic routing algorithm called 2D-HERT. In [100] a hybrid all optical network architecture called Iris is proposed that integrates *(i)* a dielectric antenna array based broadcast subnetwork transporting latency-critical short messages and *(ii)* a circuit-switched mesh subnetwork carrying throughput-bound workloads. Both broadcast and circuit-switched subnetworks operate in tandem to provide low latency, high-throughput and balanced power performance tradeoffs.

Unlike the works discussed above, our proposed *METEOR* hybrid nanophotonic-electric communication architecture has a configurable photonic ring that augments a traditional 2D all-electrical mesh NoC. The photonic waveguides in our architecture are configured as a combination of SWMR (Single Writer Multiple Readers) and MWMR (Multiple Writers Multiple Readers) to achieve cost savings. Waveguide crossings in photonic torus and some crossbar architectures can lead to losses that increase dissipated power, which is avoided when utilizing a ring topology. Another important issue is the latency for setting up the transfers and sending acknowledgements via the electrical NoC in some of these architectures, which can dramatically reduce performance and increase energy consumption according to our studies. Our architecture utilizes the much faster on-chip photonic infrastructure for path setup and flow control. The proposed architecture is also much simpler than complex crossbar and torus architectures resulting in a reduced photonic path complexity while still providing significant opportunity for improvements over traditional, all-electrical NoCs. Finally, the support for an adaptive PRI (Photonic Region of Influence) enables finer granularity trade-offs while balancing traffic between the electrical and photonic interconnects.

Most of the work on designing NoCs has focused on optimization for an application with only a single operating mode (use case), e.g. [49] [109] [110]. These techniques lead to sub-optimal designs for today's applications with multiple use cases [111]. Several works [112] [113] [114] [115] [116] [117] [118] have explored dynamic reconfiguration to adapt to changing traffic patterns for a single use case application, using dynamic voltage scaling/dynamic frequency scaling (DVS/DFS), adaptive routing schemes, and adaptive arbitration to improve performance and power dissipation. Only recently have a few approaches started to focus on designing and optimizing NoCs to meet performance constraints of multiple use-cases. Murali et al. [119] [120]

proposed using DVS/DFS, adaptive routing, and adaptive time division multiple access (TDMA) slot allocation to reduce NoC power dissipation. Hansson et al. [111] [121] described an optimization technique to reduce switching time between use cases on a mesh NoC fabric. However, none of these works have explored runtime reconfiguration for hybrid nanophotonic-electric NoC architectures to optimize power dissipation. Our proposed *UC-PHOTON* architecture enables several new techniques for runtime optimization, in addition to existing techniques such as DVS/DFS, clock gating, adaptive TDMA slot allocation, and adaptive arbitration, for multi-use case applications.

Recent advances in silicon photonics have led to the development of fabrication technologies to stack optical devices in multiple layers [122] in 3D ICs. In [104] a multi-layer hybrid nanophotonic-electric 3D NoC architecture was proposed based on a 3D crossbar topology. Our proposed *OPAL* architecture, in contrast, was based on a hybrid 3D multi-ring/mesh topology to improve performance scalability for emerging CMPs with hundreds of cores. Unlike [104], our architecture considers runtime reconfiguration of the electrical and photonic networks with the goal of significantly reducing communication power dissipation.

Recently, Xue et al. [42] presented a novel intra-chip single-hop free-space nonophotonic interconnect based on VCSELs (vertical cavity surface emitting lasers) along with an algorithm to address challenges of free-space point-to-point optical link collision. Abousamra el al. [43] extended the work from [42] to create a two-hop free-space network that significantly reduced VCSELs required for the on-chip network. Our proposed *HELIX* and *3D-HELIX* hybrid nanophotonic-electric NoC fabrics based on free-space photonic links differ compared to these previously proposed architectures by *(i)* utilizing CMOS compatible energy-efficient MQW modulators and detectors coupled to an external laser instead of bandwidth-limited and failure-

prone VCSEL devices for on-chip free space optical interconnects (FSOI); *(ii)* integrating a novel FSOI hybrid routing and flow control scheme that can be configured either for single or multi-hop communication; and *(iii)* incorporating a synthesizable FSOI collision detection and mitigation mechanism that can be dynamically configured through the serializer/deserializer modules.

2.3 NOC SYNTHESIS

Current research on application-specific NoC synthesis [45] [46] [47] [50] [51] [52] has mainly focused on electrical NoCs. For instance, Murali et al. [45] presented a floorplan aware synthesis technique that considers wiring complexity of the NoC during topology synthesis along with min-cut partitioning to allocate switches to groups of custom cores and minimize NoC power. Srinivasan et al. [46] presented a low complexity genetic algorithm based approach to synthesize a low power custom NoC topology. Chatha et al. [47] presented synthesis techniques for an application-specific NoC that employed integer linear programming (ILP) and min-cut/flow algorithms as well as node-weighted Steiner trees to obtain shortest paths. One limitation of these approaches is that they target single applications, which is increasingly impractical for today's multi-programmed workloads. Other techniques aimed at regular NoC synthesis have mainly focused on mapping uniform size cores and their communication flows on regular mesh topologies [50] [51] [52]. For instance, Ascia et al. [50] use a genetic algorithm approach to map cores to minimize communication power in a mesh NoC. Kapadia et al. [52] proposed a heuristic approach to enable a voltage island-aware low-power mapping of cores and communication routes on a regular mesh NoC. None of these synthesis approaches for regular or irregular NoCs has focused on synthesizing hybrid nanophotonic-electric NoCs.

3 *METEOR:* HYBRID PHOTONIC RING-MESH NOC FOR MULTICORE ARCHITECTURES

In this chapter we propose *METEOR*, a novel hybrid nanophotonic-electric ring-mesh NoC that utilizes configurable photonic ring waveguides coupled with a traditional 2D electrical mesh NoC. We present many new concepts that enhance the performance, latency and power compared to the previously proposed NoC architectures with minimum area usage in silicon, metal as well as nanophotonic layers.

3.1 SYSTEM LEVEL ARCHITECTURE

Our novel hybrid ring-mesh electro-photonic NoC fabric *METEOR* for emerging CMPs is based on advances in nanoscale silicon photonics with commercial CMOS manufacturing technology. Our proposed fabric consists of a photonic ring waveguide that acts as a global communication channel and complements a more traditional 2D electrical NoC fabric. This hybrid communication architecture utilizes electrical and photonic paths simultaneously to improve the performance-per-watt characteristics of a CMP. We explore different architectural *configurations* of our hybrid photonic NoC fabric by considering *(i)* varying levels of electrical to photonic communication connectivity, *(ii)* multiple degrees of communication serialization, and *(iii)* different levels of photonic wavelength division multiplexing. These configurations enable interesting tradeoffs between performance and power consumption in the proposed architecture. Our experimental results indicate significant potential for *METEOR* as it can provide about 5× reduction in power consumption and improvements in throughput and access latencies, compared to traditional electrical 2D mesh and torus NoCs. Our proposed *METEOR* fabric also demonstrates lower photonic layer area cost, power consumption, and energy-delay

product, while maintaining competitive communication latency and throughput compared to previously proposed hybrid photonic NoC fabrics, such as *(i)* the hybrid photonic torus [102], *(ii)* the all-optical Corona crossbar [40], and *(iii)* the hybrid hierarchical Firefly crossbar [37]. *METEOR* consists of concentric ring waveguides and photonic components on a dedicated photonic layer, interfacing with a traditional 2D electrical mesh NoC using interfaces that comprise driver buffers, TIA amplifiers, circuitry for clock synchronization and recovery, as well as serialization and de-serialization. Data flits are transferred in the network using wormhole switching, with flit width = *128* or *256*. There are two types of electrical layer routers used in our proposed architecture: *(i)* four stage pipelined electrical mesh routers (with the following pipeline stages: buffer write/route computation, region validation/switch allocation, switch traversal, link traversal) that have *5 I/O* ports *(N, S, E, W, local core)* with the exception of the boundary routers that have fewer I/O ports, and *(ii)* gateway interface routers that are also four-stage pipelined but have six *I/O* ports *(N, S, E, W, local core, photonic link interface)* and are responsible for sending/receiving flits to/from photonic interconnects in the photonic layer. Both types of routers have an input and output queued crossbar with a 4-flit buffer on each input/output port, with the exception of the photonic ports in gateway interface routers that use double buffering to more effectively cope with the higher photonic path throughput.

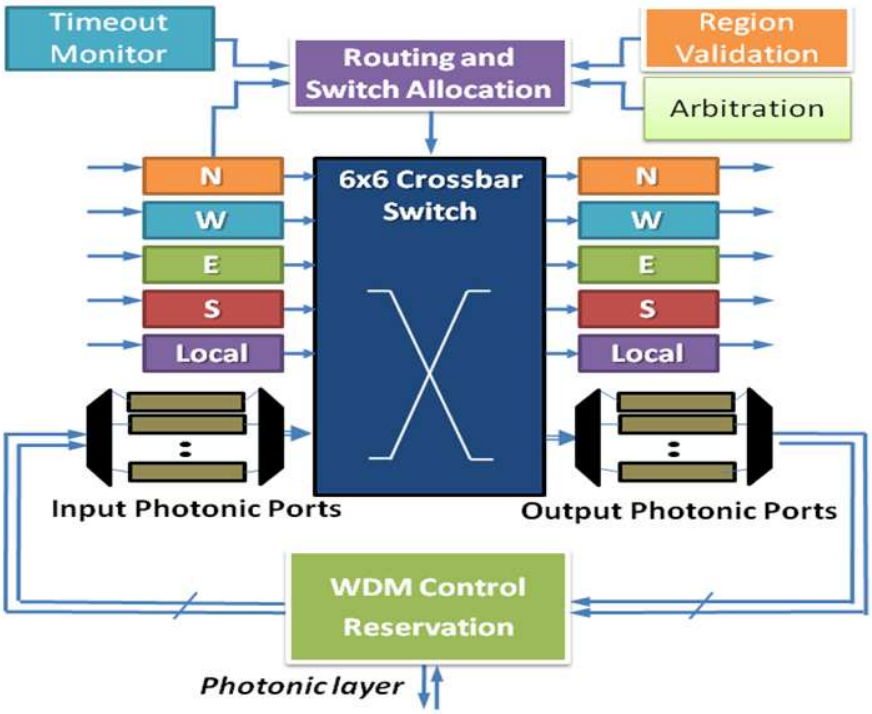
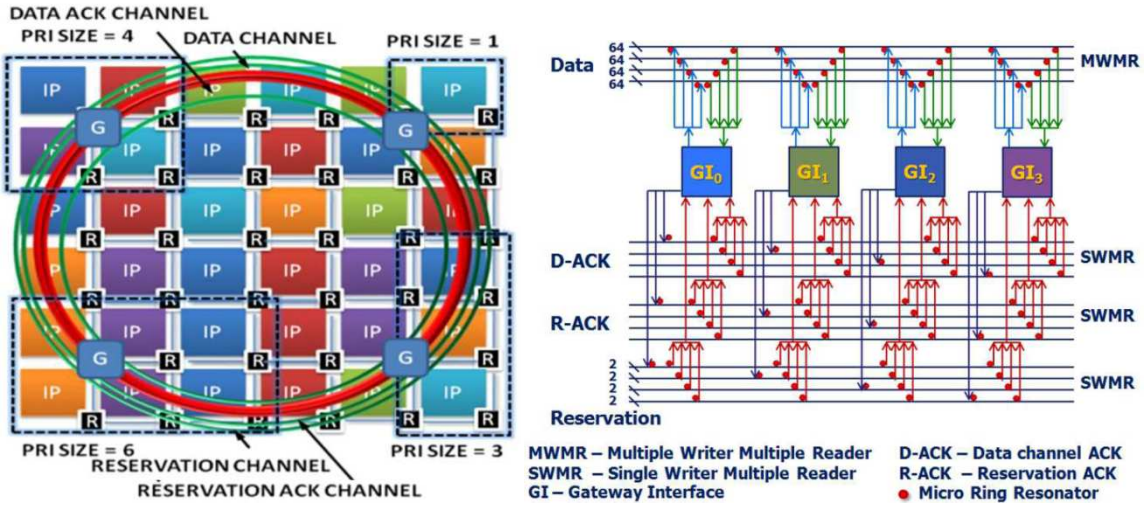


Figure 13 Gateway interface electrical router architecture

Figure 13 shows the high level architecture of a gateway interface router. Note the additional ports for interfacing with the photonic rings. Because each port has access to □ / *n* wavelengths for transmission, we have □ / *n* (double) buffers for sending data. Although it is theoretically possible to have *(1-n)*×□ / *n* data flows received at a gateway interface, we restrict the number of received flows (and hence receive buffers) to □ / *n* to maintain symmetry and reduce cost. All of the photonic ports are connected to a *WDM control* module that controls wavelength assignment to different traffic flows, to enable *WDM* for high bandwidth photonic communication. To reduce the overhead on router complexity, only a very few number of routers (four in our initial baseline configuration) are chosen as gateway interface routers. The extent of

communication over the photonic waveguide is controlled by a parameterizable photonic region of influence (PRI), described next.



(a) **(b)**

Figure 14 (a) Photonic regions of influence (PRI) (b) SWMR reservation channels and MWMR data channels

3.2 PHOTONIC REGIONS OF INFLUENCE (PRI)

To keep implementation costs low, we would like to restrict the number of gateway interfaces in the *METEOR* architecture. However, this may cause the amount of photonic path utilization to reduce as the size of the CMP and number of cores on a chip increase. To ensure appropriate scaling and utilization, we proposed a parameterizable photonic region of influence (PRI), which refers to the number of cores around the gateway interface that can utilize the photonic path for communication. For smaller sized systems (e.g., 3×3 CMPs), limiting the number of cores interfacing with each gateway interface to one (i.e., region of size 1) may be sufficient to offload a majority of the global communication from the electrical network. However for more complex systems (e.g., 8×8 CMPs) a larger region size may be more

appropriate. Figure 14 (a) shows an 8×8 CMP with different PRI sizes at the four gateway interfaces. If a router falls under the region of photonic influence, it is modified to additionally consider the photonic path for global communication. Note that while the sizes for the regions are shown as different at each gateway interface Figure 14 (a), this is for illustration purposes only, and in practice we assume a fixed sized PRI for all gateway interfaces. Also another key thing to note is that all cores need not to be part of PRI regions and cores that are outside of PRI regions communicate only through electrical network. Our experiments explore the impact of varying the PRI size on overall performance and power consumption. Details of routing and flow control in *METEOR* are presented next.

3.3 ROUTING AND FLOW CONTROL

To route flits in *METEOR*, an XY dimension order routing scheme is used in the electrical NoC, and a modified PRI-aware XY routing scheme is employed for selective data transmission through the photonic links. Communicating cores lying within the same PRI region communicate using the electrical NoC (intra-PRI transfers). Cores that need to communicate and reside in different PRI regions communicate using the photonic paths (inter-PRI transfers), provided they satisfy two criteria: *(i)* the size of the data to be transferred is above a user-defined threshold M_{th}, and *(ii)* the number of hops from the source core to its local PRI gateway interface is less than the number of hops to its destination core. The inter-PRI routing process is implemented by adding an *'inter-PRI'* bit within the header flit. Changing PRI region size involves updating region boundary coordinate units in *'region validation'* tables of the NoC routers (Figure 13). The header flit *'inter-PRI'* bit is set according to the *'region validation'* table values. For inter-PRI communication, if the *'inter-PRI'* bit is set then flits are routed towards the source gateway

interface that falls within PRI and then routed towards the destination gateway interface through the nanophotonic waveguides.

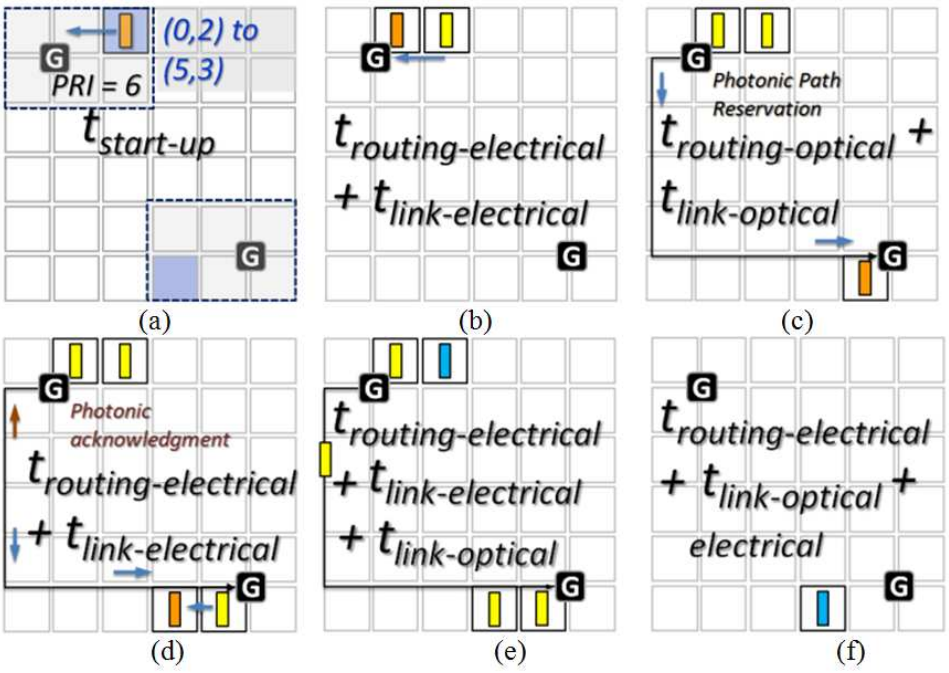


Figure 15 Flit life cycle during inter-PRI wormhole routing path (a) processor initiates communication (b) header flit (orange) is routed to nearest gateway (c) communication of header through photonic path (d) header flit completes path reservation and reaches destination (e) data flit (yellow) transmission continues (f) path is dismantled by tail flit (blue)

Our implementation is constrained to a single gateway interface per PRI to minimize implementation complexity. Transfers between cores lying outside PRI regions occur via the electrical network using XY routing. Network interfaces (NIs) ensure that header flits contain coordinates of the source and destination of the packet being injected into the NoC, as well as a flag indicating that the message size is large enough to potentially traverse a photonic path (for inter-PRI transfers). In case of overlapped PRI regions, the region validation table includes multiple entries for the nodes covered by multiple PRI regions and flits are always routed towards the nearest gateway interface. If the distance from the source node to two or more

gateway interfaces is the same, then flits are routed towards the gateway interface that is closer to the destination gateway interface. Nodes lying within overlapped regions are considered part of PRI regions that contribute to the overlap. Communication between these nodes and nodes within any of the overlapping PRI regions is treated as intra-PRI data transfers.

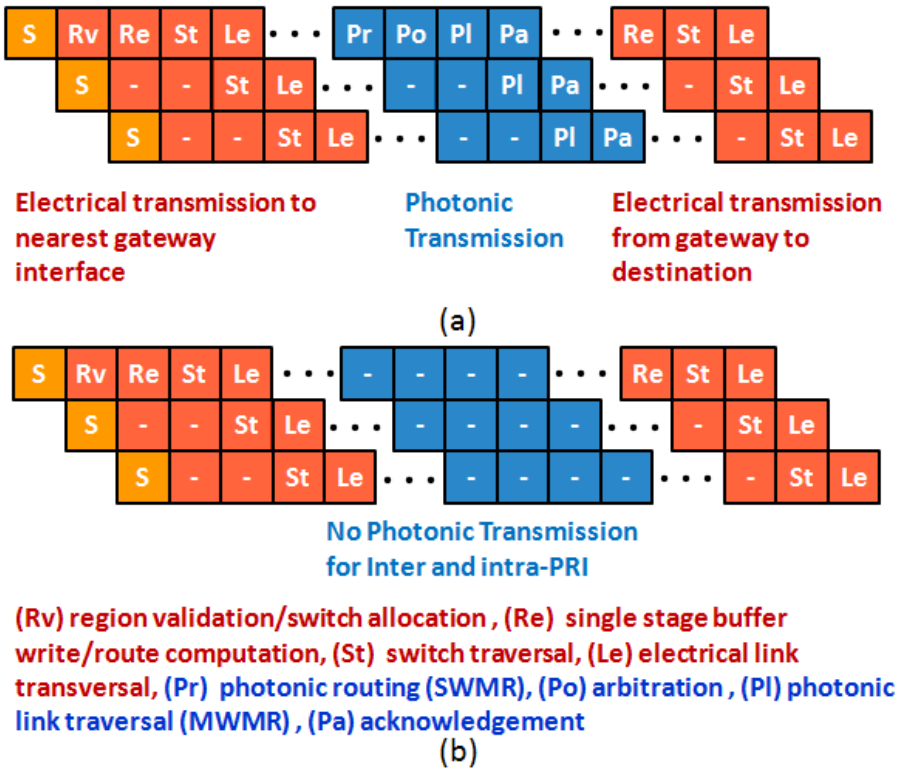


Figure 16 Head, body and tail flit routing pipeline for (a) inter PRI transfer (b) intra-PRI and non-PRI transfer, dots represent multiple data flit transfers

The photonic waveguides in *METEOR* are logically partitioned into four channels: reservation, reservation acknowledge, data, and data acknowledge as shown in Figure 14 (a) and (b), achieving more than 12 and 24 TB/s bandwidth with 128 and 256 data waveguides respectively. In order to reserve a photonic path for a data transfer, *METEOR* utilizes a Single Write Multiple Read (SWMR) configuration on dedicated reservation channel waveguides.

SWMR is less sensitive to the laser static power and modulator insertion loss compared to MWSR [123]. Each gateway interface has a subset of λ/n wavelengths (microresonator modulators) available for transmission, where λ is the total number of wavelengths available from the multi-wavelength laser and n is the number of gateway interfaces. Every gateway interface must be able to receive $(n-1)\times\lambda/n$ wavelengths (from the rest of the gateway interfaces), each with a separate microring resonator receiver. A source gateway interface uses one of its available wavelengths λ_t to multicast the destination ID via the reservation channel to other gateway interfaces. Each gateway interface has $\lceil log\ (n-1) \rceil$ dedicated SWMR reservation photonic waveguides that it writes the destination ID to, after which the other gateway interfaces read the request. Only the intended destination gateway interface accepts the request, while the others ignore it. As each gateway interface has a dedicated set of λ/n wavelengths allocated to it, the destination can determine the source of the request, without the sender needing to send its ID with the multicast.

If the request can be serviced by the available wavelength and buffer resources at the destination, a reservation acknowledgement is sent back via the reservation ACK channel on an available wavelength. The reservation ACK channel also has a SWMR configuration, but a single waveguide per gateway interface is sufficient to indicate the success or failure of the request. Once the photonic path has been reserved in this manner, data transfer proceeds on the data channel, which has a low cost Multiple Writers Multiple Readers (MWMR) configuration. Figure 15 illustrates this process, including the latencies involved for an Inter-PRI communication pattern. As flits are routed through the nearest gateway interface, global communication power consumption is significantly lowered and the electrical network bandwidth availability is increased, enabling a win-win scenario. Our experimental results

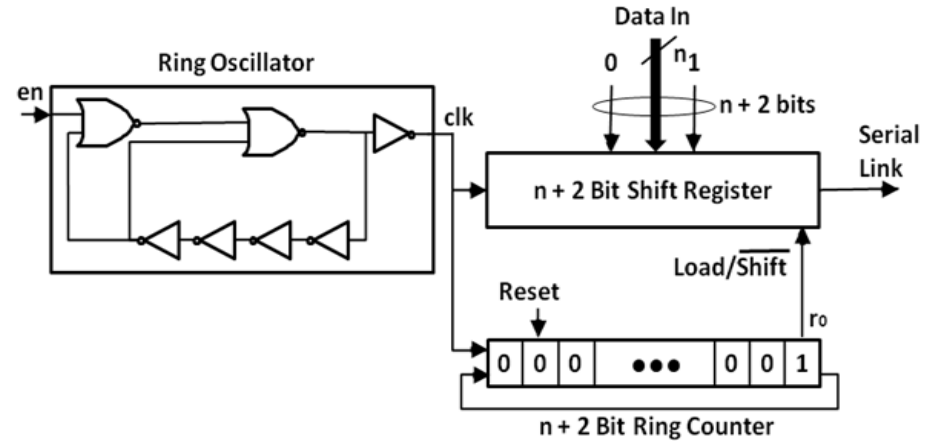
quantitatively demonstrate these benefits in detail. The default number of data channel waveguides is equal to the chosen flit width to enable consistency and ease integration with the electrical network interface, where data links are also equal to the flit width, thus allowing more consistent design for the network routers. The same wavelength λ_t used for the reservation phase is used by the source to send data on. The destination gateway interface tunes one of its available microring resonators to receive data from the sender on that wavelength after the reservation phase. Once data transmission has completed, an acknowledgement is sent back from the destination to the source gateway interface via a data ACK channel that also has a SWMR configuration with a single waveguides per gateway interface to indicate if the data transfer completed successfully or failed. The advantage of having a fully photonic path setup and ACK/NACK flow control in *METEOR* is that it avoids using the high latency electrical network, as is proposed with some other approaches [38] [101] [102] [106]. As our analysis will show, the novel combination of SWMR reservation and MWMR data channel schemes in *METEOR* can provide a major advantage towards mitigating power and latency bottlenecks. Allowing gateway interfaces to request for access to the photonic paths whenever data is available is also more efficient than using a token ring scheme, which can suffer from low throughput and high latencies, especially under low traffic conditions. Note that acknowledgements are essential during photonic transfers because data transfers through nanophotonic waveguides can fail due to factors such as crosstalk and low signal to noise ratio. Some recent research efforts have explored techniques to mitigate channel noise. As an example, [124] highlighted the detrimental effect of crosstalk in on-chip photonic waveguides and presented a technique to overcome this crosstalk.

Figure 16 depicts our pipelined data packet transfer process, where *S* is the clock cycle to generate a data packet within a processor core. The data transmission proceeds to the nearest gateway with region validation/switch allocation R_V, single stage buffer write/route computation R_E, switch traversal S_T and electrical link transversal L_E. Dots indicate multiple hops in some cases to generalize this diagram. The SWMR configuration broadcasts the reservation flit followed by photonic routing P_R and arbitration P_O. The MWMR photonic link traversal transfers data P_L followed by acknowledgement P_A. The inter-PRI gateway interface process proceeds through all of the above cycles for the header flit. Subsequent data flits in the electrical path can skip region validation R_V and route computation R_E stages. Similarly, subsequent data flits in the photonic path can skip photonic routing P_R and arbitration P_O (Figure 16 (a)). Intra-PRI and non-PRI transfers do not utilize photonic transmission and thus the photonic steps are skipped for these types of transfers Figure 16(b)).

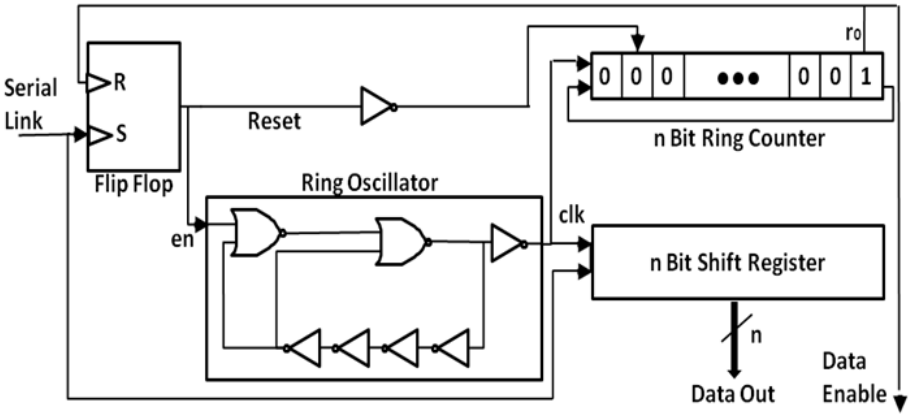
3.4 DEADLOCK RECOVERY

While XY routing has been proven to be deadlock-free for mesh-like regular NoCs (as no channel dependency cycles can be formed between dimensions), the modifications made to this routing scheme to accommodate photonic transfers in *METEOR* may end up creating deadlock conditions. We extensively studied deadlocks in the proposed architecture when packets traverse the photonic ring paths. To overcome a potential deadlock, we arrived at using low overhead timeout flits sporadically interleaved with the flits for the long data messages traversing the photonic paths. This is a form of regressive deadlock recovery [125]. If a timeout flit reaches a router where flits are blocked, a timeout monitor module in the router can detect a timeout event and recognize potential cases where flits are blocked due to deadlock, and drop the blocked flits,

while sending a NACK signal in the reverse direction to indicate the flits being dropped. This allows the system to unblock and recover from potential deadlock. While the method has the overhead of the additional flits in long messages intended for photonic links and monitoring module in the routers, this is still simpler than other potential deadlock resolution alternatives proposed e.g., a common technique is to keep extra escape channels in every router and draining deadlocked packets through the escape channels until the deadlock condition clears [125].



(a)



(b)

Figure 17 Serialization scheme for gateway interface (a) serializer, (b) de-serializer

3.5 COMMUNICATION SERIALIZATION

While the *METEOR* baseline configuration does not use serialization, we study serialization as a potential design alternative to reduce area and power. Serialization of electrical communication links has been widely used in the past to reduce wiring congestion, lower power consumption (by reducing link switching and buffer resources), and improve performance (by reducing crosstalk) [126] [127] [128]. Typically serialization in the electrical realm allows increasing frequency in serialized copper links to make them faster compared to parallel links. This is not the case in our architecture where serialization is mainly a means to shut down a subset of photonic components and save power. As reducing power consumption is a critical design goal in future CMPs, we proposed using serialization at the gateway interfaces, to reduce the number of photonic components (waveguides, buffers, transmitters/receivers), and consequently reduce area and complexity on the photonic layer as well as lower the power consumption. In our architecture, we make use of a shift register based serialization scheme, similar to [129] [130] [131]. A single serial line is used to communicate both data and control signals between the source and destination nodes. A frame of data transmitted on the serial line using this scheme consists of *n+2* bits, which includes a start bit *(1)*, *n* bits of data, and a stop bit *(0)*. Figure 17 (a) shows the block diagram of the transmitter (or serializer) at the source. When a word is to be transferred, the ring oscillator is enabled and it generates a local clock signal that can oscillate above 2 GHz to provide high transmission bandwidth. At the first positive edge of this clock, an *n+2* bit data frame is loaded in the shift register. In the next *n+1* cycles, the shift register shifts out the data frame bit by bit. The stop bit is eventually transferred on the serial line after *n+2* cycles, and *r0* becomes *1*. At this time, if the transmission buffer is empty, the ring oscillator and shift registers are disabled, and the serial line goes into its idle state. Otherwise, the

next data word is loaded into the shift register and data transmission continues without interruption. Table 1 shows how serialization degree impacts performance of the photonic links.

Table 1 Serialization link bandwidth

Serialization Degree	*WDM*	*Waveguides*	*Electrical Flit-Width*	*Photonic BW [TB/s]*
1:1	32	128	128	12
2:1	32	64	128	6
4:1	32	32	128	3
8:1	32	16	128	1.5
1:1	32	256	256	24
2:1	32	128	256	12
4:1	32	64	256	6
8:1	32	32	256	3

Figure 17 (b) shows the block diagram of the receiver (or deserializer) at the destination. An *R-S* flip-flop is activated when a low-to-high transition is detected on the input serial line (the *low* corresponds to the stop bit of the previous frame, while the *high* corresponds to the start bit of the current frame). After being activated, the flip-flop enables the receiver ring oscillator (which has a circuit similar to the transmitter ring oscillator) and the ring counter. The *n*-bit data word is read bit by bit from the serial line into a shift register, in the next *n* clock cycles. Thus, after n clock cycles, the n bit data will be available on the parallel output lines, while the least significant bit output of the ring counter *(r0)* becomes *1* to indicate data word availability at the output. With the assertion of *r0*, the *R-S* flip-flop is also reset, disabling the ring oscillator. At this point the receiver is ready to start receiving the next data frame. In case of a slight mismatch between the transmitter and receiver ring oscillator frequencies, correct operation can be ensured by adding a small delay in the clock path of the receiver shift register. The preceding discussion

assumed *n:1* serialization, where *n* data bits are transmitted on one serial line (i.e., a serialization degree of n). If wider links are used, this scheme can be easily extended. For instance, consider the scenario where *4n* data bits need to be transmitted on four serial lines. In such a case, the number of shift registers in the transmitter must be increased from *1* to *4*. However the control circuitry (flip-flop, ring oscillator, ring counter) can be reused among the multiple shift registers and remains unchanged. At the destination, every serial line has a separate receiver to eliminate jitter and mismatch between parallel lines.

3.6 EXPERIMENTAL RESULTS

Photonic waveguides provide faster signal propagation compared to electrical interconnects because they do not suffer from RLC impedances. But in order to exploit the propagation speed advantage of photonic interconnects, electrical signals must be converted into light and then back into an electrical signal. This process requires a performance and power overhead that must be taken into account for an accurate analysis. In this section, we present experimental results to evaluate our proposed *METEOR* communication fabric that combines electrical and photonic interconnects. The first two subsections present the simulation setup and details of the estimation models used. The subsequent subsections present our experimental results, including comparisons with other hybrid photonic NoCs.

3.7 SIMULATION SETUP

For our experimental studies, the hybrid electro-photonic *METEOR* architecture was modeled at the cycle accurate granularity by extensively modifying our in-house cycle accurate SystemC-based NoC simulator that was derived from the open-source Nirgam [132] and Noxim [133] NoC simulators.

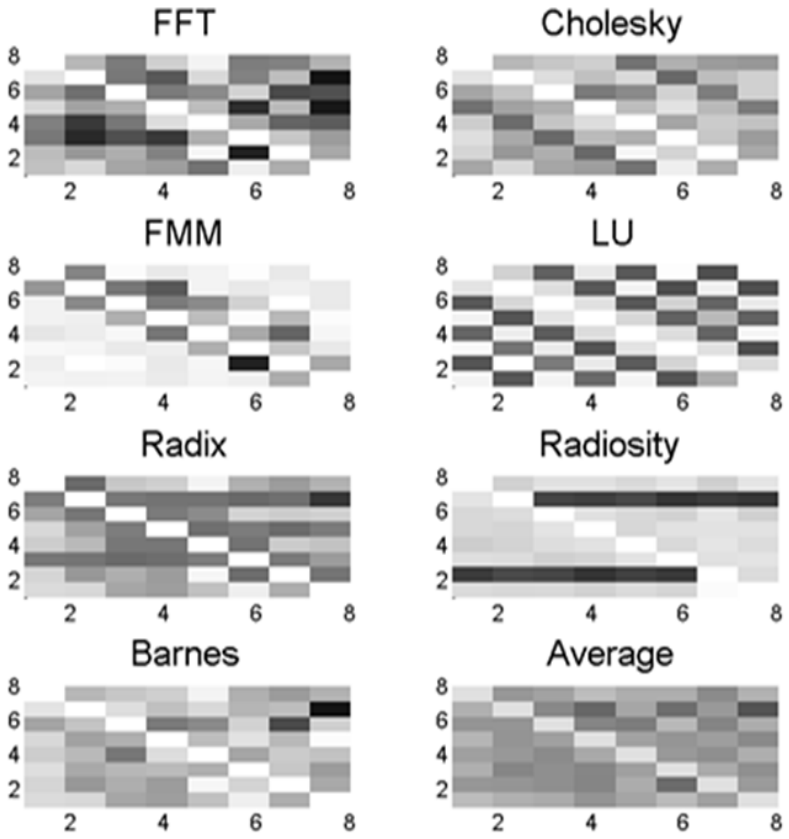


Figure 18 *SPLASH-2* implementation traffic maps for 8x8 CMP

We targeted a 32 nm process technology, and assumed a 400 mm^2 CMP die area. We used a high level floorplanner [134] to compute wire lengths for various NoC configurations, and obtain accurate power and performance estimates. The operating frequency of the photonic ring is estimated by calculating the time needed for light to travel from any node to the farthest node on the (unidirectional) ring, so that data can be transmitted to all nodes in one cycle. We assume the presence of last level cache banks and I/O controllers around the periphery of the chip, which results in a photonic ring with a diameter that is smaller (*~14 mm*) than the chip edge width. Through geometric calculations for this ring and assuming a refractive index of 3 for the SOI

waveguide, we were able to clock the photonic ring (and the communication network) at a frequency of 2.3 GHz.

We used benchmarks with different synthetic traffic profiles (*Hotspot, Bitwise, Shuffle, Transpose, Butterfly, Uniform Random*) to explore architectural performance under diverse traffic conditions. For comparisons with other hybrid photonic NoCs, we additionally implemented seven benchmarks from the *SPLASH-2* suite [135] (*Cholesky, FFT, Fmm, Lu, Radiosity, Radix, Barnes*) and these were used to load traffic on the communication fabric. Figure 18 shows the traffic distribution for the *SPLASH-2* benchmarks implemented on an *8×8* CMP. Each cell represents a core, with lighter colored cores sending/receiving fewer packets than darker colored cores. We also performed comparisons using the *PARSEC* and *NAS* benchmark suites [136] [137].

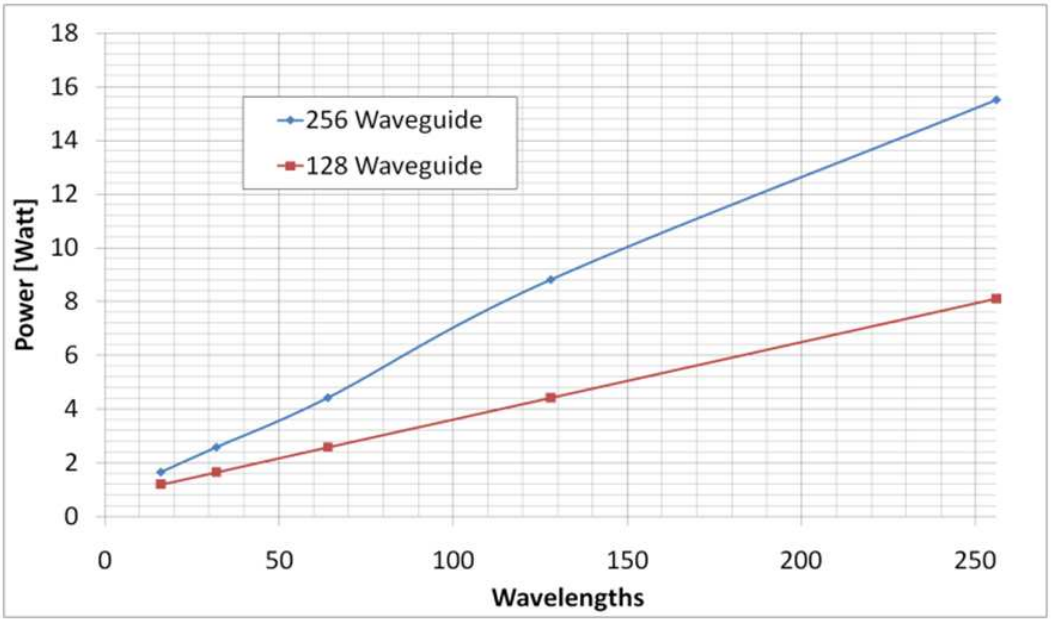


Figure 19 *METEOR* laser power for different degrees of WDM

Table 2 Delay and power of photonic components

Component	Delay	DDE	SP	TTE
Modulator driver	*9.5 ps*	20 *fJ/bit*	*5 μW*	16 *fJ/bit/heater*
Modulator	*3.1 ps*			
Waveguide	*15.4 ps/mm*	-	-	-
Photo Detector	*0.22 ps*	20 *fJ/bit*	*5 μW*	16 *fJ/bit/heater*
Receiver	*4.0 ps*			

Note: Delay and power consumption for METEOR elements (32nm) DDE = Data traffic dependent energy, SP = Static power, TTE = Thermal tuning energy (20K temperature range) [36] [138] [139]

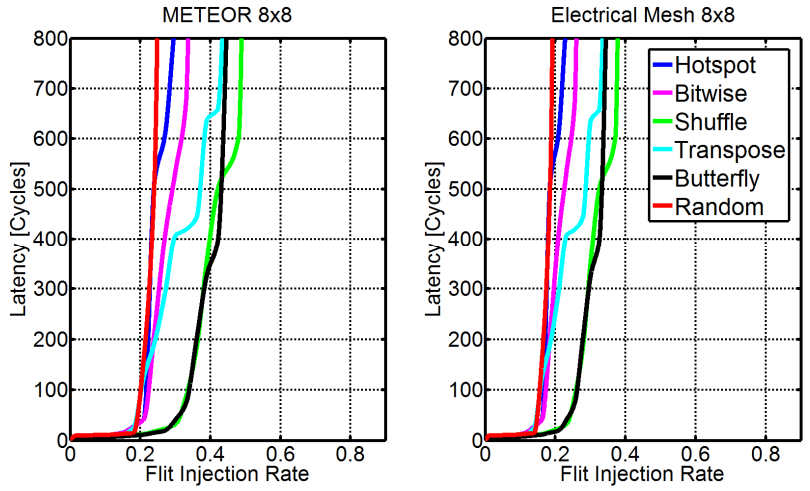
3.7.1 PERFORMANCE AND POWER ESTIMATION MODELS

To obtain the delay for the photonic waveguide and other photonic components, we used results from [89] [140] for the 32 nm process technology node. The delay of an optimally repeated and sized electrical (Cu) wire at 32 nm was assumed to be 42 ps/mm [29]. The power consumed in *METEOR* can be divided into two parts: the power consumed in the electrical network and the power consumed in the photonic components. The static and dynamic power consumption of electrical routers in this work is based on results obtained from a modified version of the Orion 2.0 simulator [141], while the power consumption on optimally repeated and sized Cu wires is obtained from the methodology in [142]. For the power consumption of various photonic components in the *METEOR* architecture, we adopt the power models from [36] [143] derived based on device level work in [138] [139]. Based on [36] [143] thermal tuners integrated at each ring in the network consume approximately $1\mu W$ heating power per Kelvin, and have a *20K* tuning range. [139] demonstrated ~20× improvements in the thermal tuning efficiency in applied tuning power. This improvement compared to a premicromachined [138] device was achieved by reducing the power required to shift the filter resonant peak across the

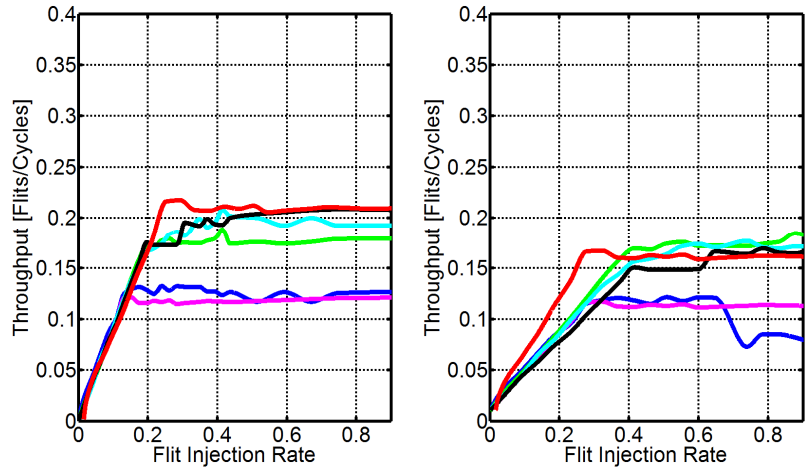
entire free spectral range. We implement microring resonator structure based on combination of device and system level work. Table 2 shows the delay and power estimation models assumed in our experiments.

Based on the processor and network bandwidth utilization, a CMP will produce temperature variations across its total surface area. This unbalanced thermal profile can require a subset of cores to operate at lower performance than others [100]. The unbalanced thermal profile also needs to be considered in the scope of microring resonator structures that can go off-resonance with thermal variations and require thermal tuning to maintain resonance and function correctly.

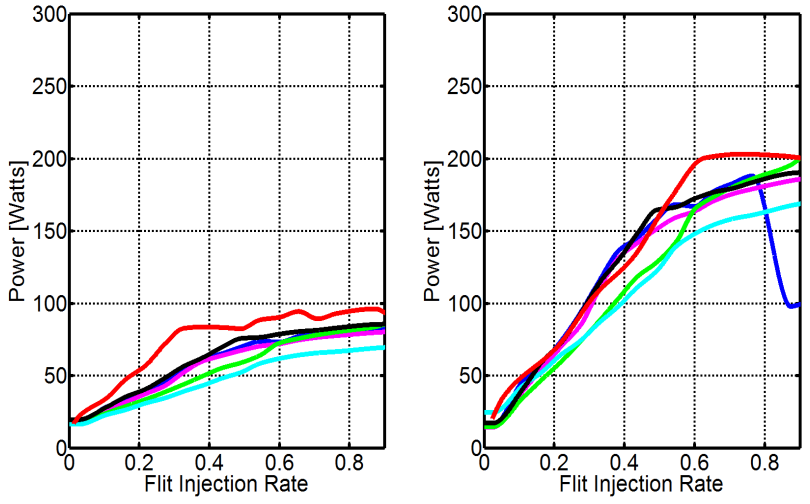
To compute laser power consumption, we calculated optical loss in components in our architecture, which sets the required optical laser power budget and correspondingly the electrical laser power. We considered per component optical losses for the coupler/splitter (1 dB) [126], non-linearity (1 dB at 30mW) [36], waveguide propagation (3 dB/cm) [144], waveguide bend (0.005 dB/90^0 bend) [144], ring modulator insertion (1 dB) [36], drop filter (1.5 dB) [145] and photodetector (0.1 dB) [36], at *30%* laser efficiency [146]. As our architecture includes concentric ring waveguides, losses due to waveguide crossings are not present. The laser intensity also needs to be compensated with the loss factor required to maintain signal integrity. Off-resonance coupling loss can occur when the optical signal passes through varying or non aligned resonance switches and modulators [147] and it can be up to 0.1 dB in our architecture. In various published literature, the sensitivity of photodetectors is assumed from $1\mu W$ [62] to $80\mu W$ [126]. In our work, we assumed the photodetector sensitivity to be $10\ \mu W$ in accordance with value from [107]. Based on the concentric ring waveguide layout, we estimated the per-



(a)



(b)



(c)

Figure 20 *METEOR* vs. electrical mesh NoC for an 8x8 NoC (a) average latency, (b) throughput, (c) power

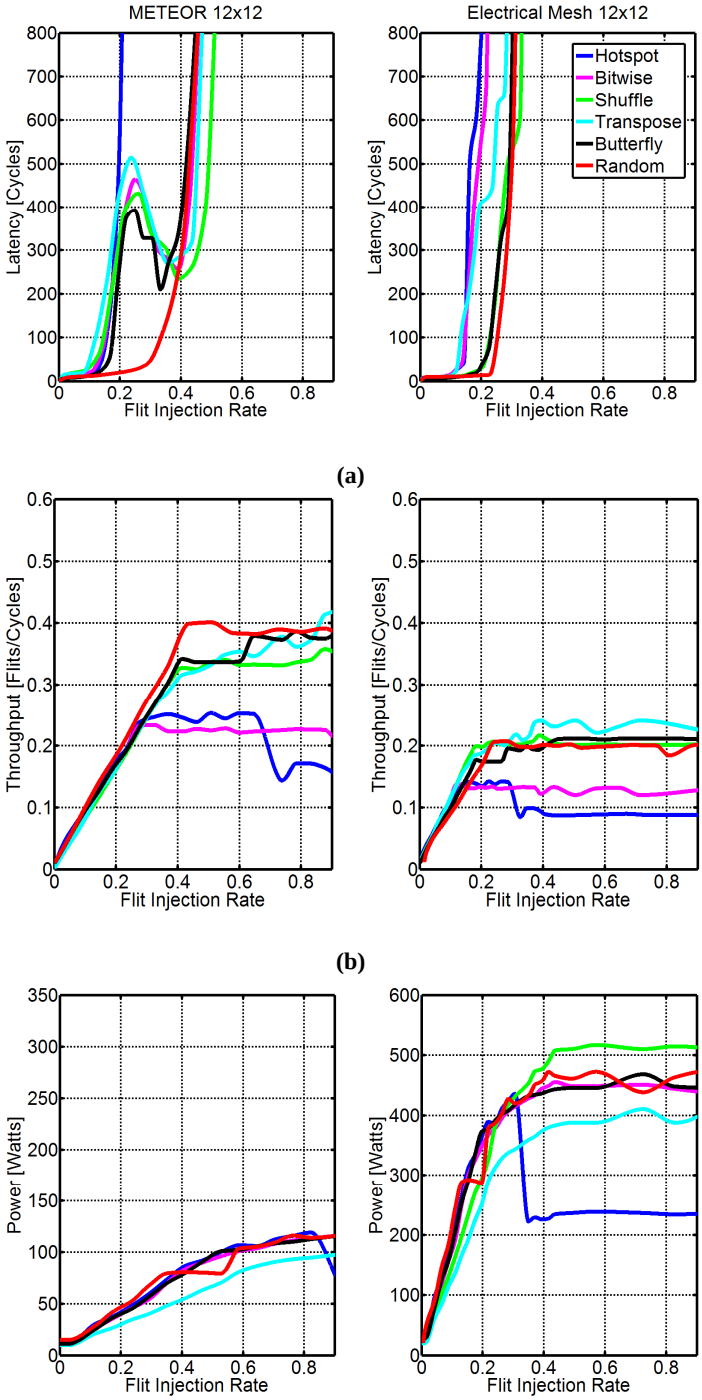


Figure 21 *METEOR* vs. electrical mesh NoC for a 12x12 NoC (a) average latency, (b) throughput, (c) power

wavelength laser power needed to offset losses and activate the farthest photodetector. Figure 19 shows the laser power in *METEOR* for various degrees of WDM (i.e., wavelengths) from 16-256 for 128 and 256 wide data path waveguide configurations of *METEOR*.

3.7.2 COMPARISON WITH ELECTRICAL MESH NOC

The first set of experiments compare our baseline hybrid photonic *METEOR* communication fabric (uplinks=4, PRI size=4, WDM degree=32) with a traditional 2D all-electrical mesh NoC, for CMPs of varying complexity. Figure 20(a)-(c) show the latency, throughput, and power consumption for a 64 core (8×8) CMP architecture with various synthetic traffic patterns. As flit injection rate from the cores increases, increased traffic congestion causes the average transfer latency from the source to the destination cores for the electrical mesh to rise rapidly. In contrast, due to the photonic ring in *METEOR* offloading a large portion of the global communication away from the electrical network, the congestion in the electrical network reduces, which results in lower average packet latency compared to the all-electrical mesh NoC. The addition of a high bandwidth photonic path also leads to a better average throughput response in *METEOR*.

This throughput however begins to saturate as the rising injection rate leads to a greater load (and thus congestion) on all the NoC components. Finally, the rate of increase in power consumption of the NoC also starts to saturate (after rapidly increasing initially) with increasing flit injection rate. *METEOR* can be seen to have a much lower power consumption compared to the all-electrical mesh NoC. Results for a larger 144 core (12x12) CMP shown in Figure 21 (a)-(c) demonstrate similar trends. However, our analysis of link loads after simulation indicated that the utilization of the photonic ring was fairly low for such large CMP sizes. One reason for this is

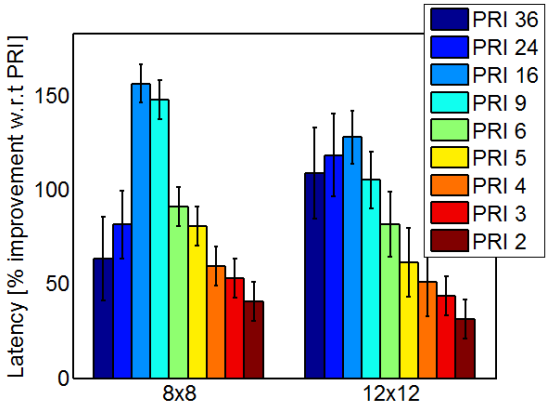
that we limited the PRI size to four, for all the experiments. As a result, the relative percentage of cores utilizing the photonic path for global communication reduces with increasing CMP size. In the next subsection, we explore the impact of varying the PRI size in *METEOR*.

In the following sections, we explore various configurations of the *METEOR* architecture starting with a baseline configuration. This baseline *METEOR* architecture is configured with the following values: uplinks=4, PRI size=4, WDM degree=32, Wavelengths=32, Serialization=1:1. Our goal is to study and quantify the impact of various configuration parameters (PRI size, number of photonic uplinks, number of wavelengths, serialization degree) on the performance and power dissipation of *METEOR*, for (8×8) and (12×12) CMPs.

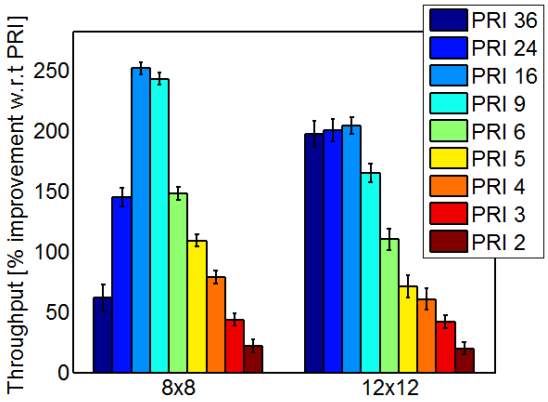
3.7.3 IMPACT OF VARYING PRI SIZE

In order to improve the utilization of the photonic path especially for large sized CMPs, we explored varying the size of the photonic regions of influence for the baseline *METEOR* configuration. Figure 22 (a)-(c) shows the percentage variation in average latency, throughput, and power consumption (relative to the base case with PRI size = 1) when the PRI size is increased from the value of 1 to 16, for the 8×8 and 12×12 CMP sizes, with results averaged over the synthetic benchmarks for brevity (trend lines on each bar show the variation).

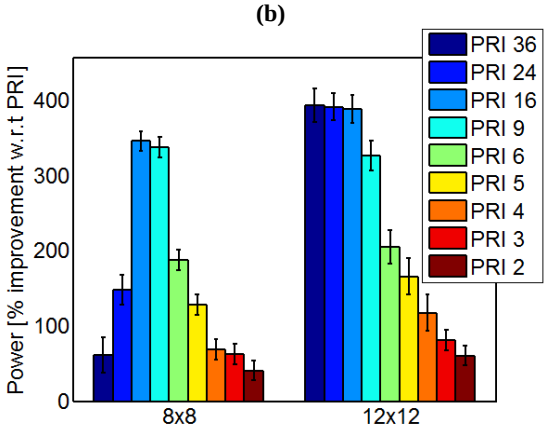
As only a single photonic ring supports the 12×12 architecture, scalability issues can be observed in performance compared to the 8×8 case. It can be seen that there is a significant performance improvement with increasing PRI size (Figure 22 (a), (b)), and the power consumption of *METEOR* also goes down (Figure 22 (c)). Note that as the PRI size increases, packets will increasingly flow through potentially several nodes in the electrical portion of the NoC, before they can utilize the photonic path. Overall, there is a decrease in average latency and power consumption, and improvement in throughput as more and more packets utilize the



(a)



(b)



(c)

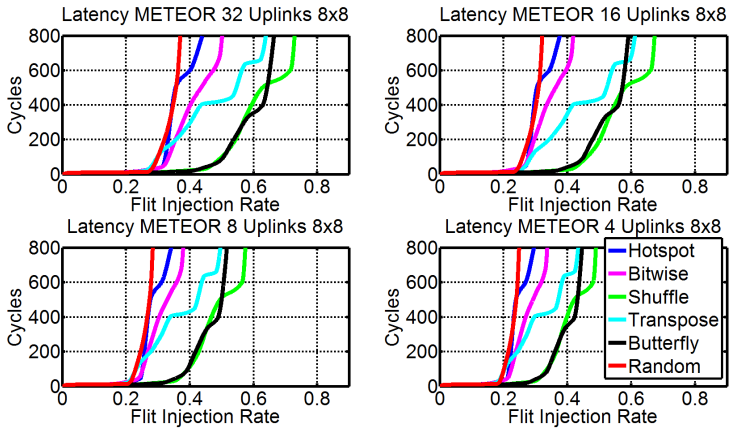
Figure 22 Relative impact of varying photonic region of influence (PRI) size in *METEOR* on (a) average latency, (b) throughput (c) power

photonic ring with increasing PRI size up to 16. Thus, increasing the PRI size is an indispensable optimization for our proposed *METEOR* communication architecture. However we also notice that as PRI size increases, the improvements saturate (especially for the 8×8 CMP case) due to the fact that more and more cores start using the photonic path with a limited number of uplinks (gateway interfaces), which creates a bottleneck at the uplinks and prevents further improvement. The increase in latency and power, and reduction in throughput on average for PRI sizes greater than 16 is due to more cores become eligible for using gateway interfaces, which increases congestion at the gateway interfaces, as well as on shared sub-paths from cores to the gateway interface where flit contention reduces performance even further. We observed increased congestion in the electrical network for such cases, and especially near the gateway interfaces, which also led to an increase in overall power dissipated due to increased power dissipation in the electrical network.

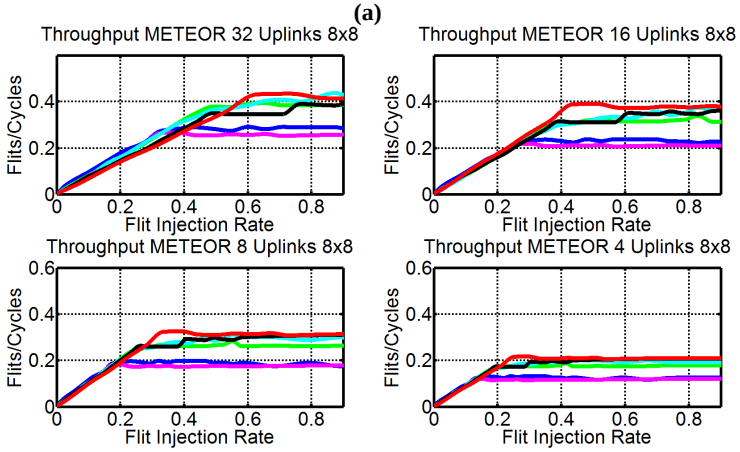
3.7.4 IMPACT OF CHANGING NUMBER OF PHOTONIC UPLINKS

To overcome the bottleneck of a limited number of up-links (i.e., gateway interfaces), we next explored the impact of changing the number of uplinks in the *METEOR* architecture and measured the performance and power dissipation for the various configurations. As the number of gateway interface routers with photonic interfaces increases, it also results in an increase in power due to O/E and E/O conversion. Increasing the number of uplinks also increases real estate usage in the silicon layer, as well as the complexity of the photonic layer.

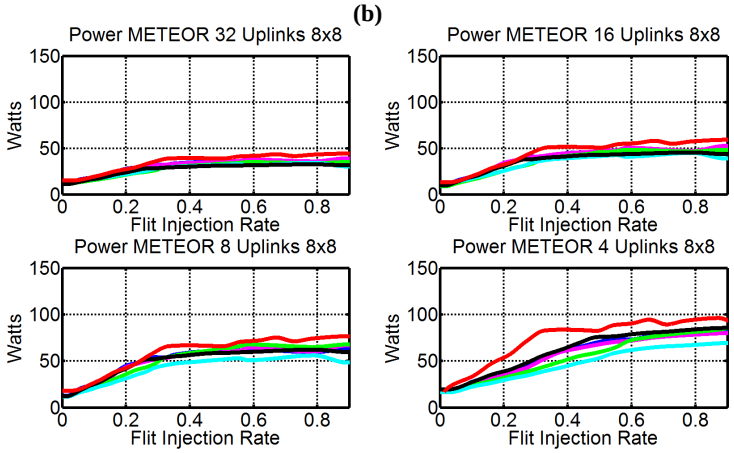
However the additional complexity of more uplinks translates into better photonic path utilization by communication flits, and can lead to lower overall power dissipation. Increasing the number of uplinks can also provide fault tolerance in case of uplink failures. Figure 23 shows



(a)

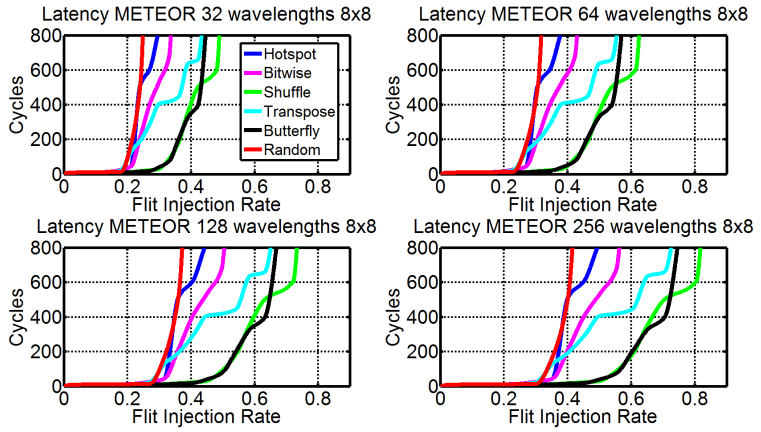


(b)

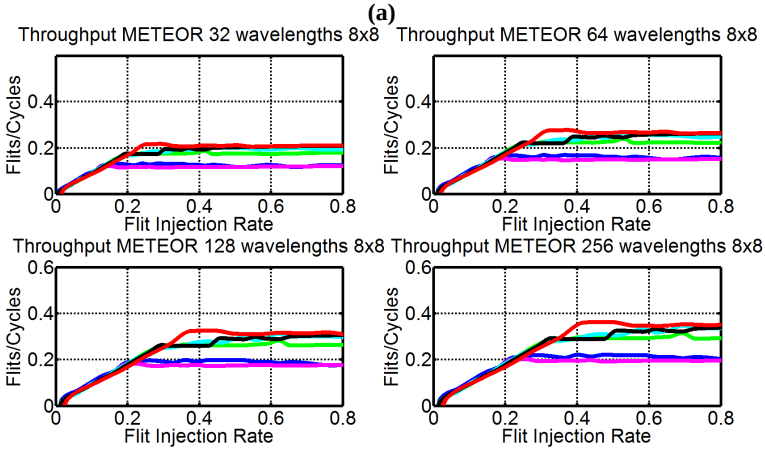


(c)

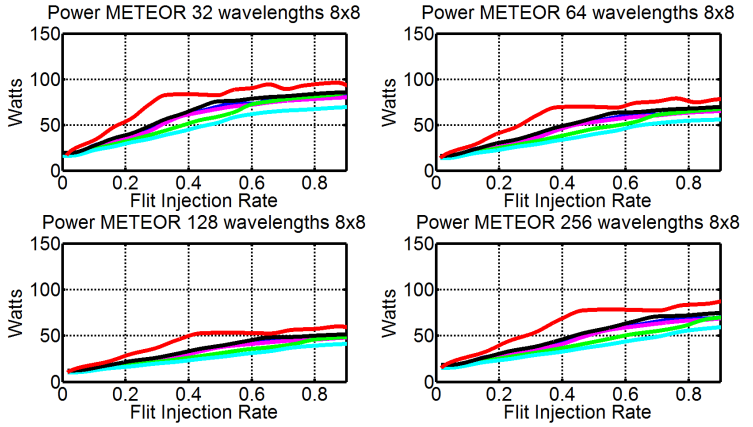
Figure 23 Impact of varying number of uplinks in *METEOR* for 8x8 NoC (a) average latency, (b) throughput (c) power consumption



(a)



(b)



(c)

Figure 24 Impact of varying number of wavelengths in *METEOR* for 8x8 NoC (a) average latency, (b) throughput (c) power consumption

results of varying the number of uplinks for an 8×8 NoC with a fixed PRI region size of 4 and WDM degree of 32. Improvements in power and performance were significant when uplinks were increased from 8 to 16. The improvements in throughput and power drop when the number of uplinks was increased from 16 to 32. This is due to overlapped PRI regions that lead to less opportunity for long distance global communication, while increasing complexity in the photonic and electrical NoC layers. Increasing the number of uplinks beyond 32 led to worse power and performance results (these results are omitted for brevity), due to much higher overheads in the electrical and photonic NoC layers.

Note that PRI size and number of gateway interfaces are closely correlated. However, given the complexity of the design space in *METEOR*, the number of all possible parameter value combinations is excessive and an exhaustive analysis is prohibitive. Therefore we chose a stepwise approach to analyze factors such as PRI size and number of gateway interfaces independently, for the sake of tractability. A more comprehensive analysis could explore the correlation between PRI size and number of gateway interfaces by analyzing all of their possible combinations, to potentially achieve improved results.

3.7.5 IMPACT OF VARYING NUMBER OF WAVELENGTHS

WDM has many practical advantages, for example, it has the ability to improve bandwidth utilization of already implemented photonic waveguides. We were interested in exploring the impact of the number of wavelengths on performance and power dissipation in our architecture.

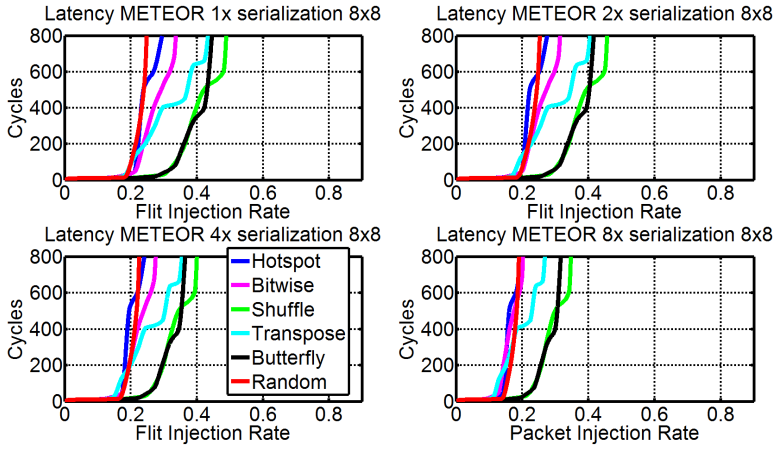
Figure 24 shows experimental results for change in average latency, throughput, and power consumption when the number of wavelengths is increased from 32 to 256 for an 8×8 CMP with 4 uplinks (gateway interfaces) and a PRI size of 4. As the degree of WDM is increased, power

consumption for the multi-mode laser, resonators, thermal tuning of each resonator and photodetectors goes up.

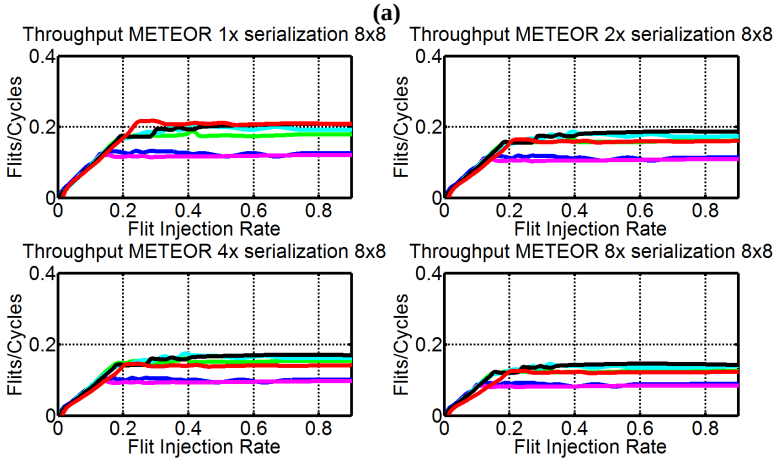
However, there is also a reduction in electrical NoC power as global communication is more quickly sent through the higher capacity photonic paths without having to be buffered for too long. The experimental results show significant improvements in average latency when the numbers of wavelengths were increased for certain benchmarks (Uniform Random, Transpose, Shuffle, Butterfly). Throughput improvements were much smaller in comparison, due to traffic bottlenecks in the electrical NoC being the limiting factor. As the number of wavelengths is increased, the performance improvements come at the cost of higher power consumption overhead (especially when the number of wavelengths is increased from 128 to 256). Therefore it is important to carefully balance performance and power needs on a per application basis when selecting the WDM degree.

3.7.6 IMPACT OF PHOTONIC SERIALIZATION

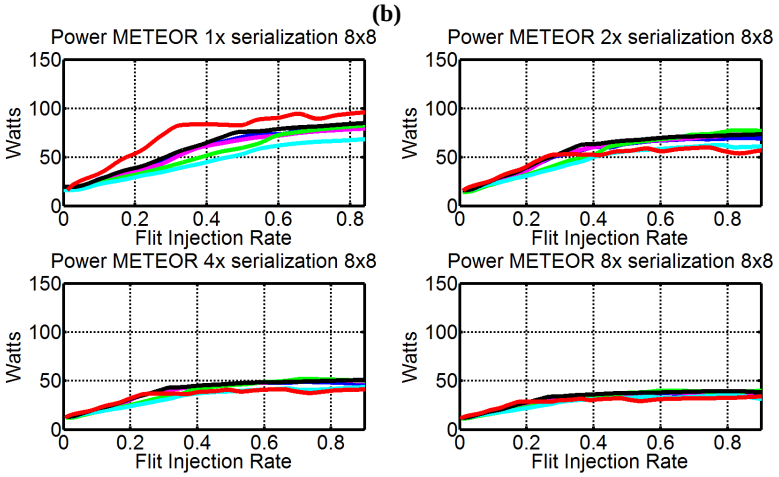
In order to further reduce power consumption, we explored using data serialization for transfers over the photonic waveguide. The goal is to minimize the E/O and O/E conversion circuitry and buffer sizes as well as switching activity, to reduce power consumption. Figure 25 (c) shows the reduction in power consumption for our *METEOR* architecture (uplinks=4, PRI size=4, WDM degree=32) as the degree of serialization is changed from the original un-serialized case (1×) to 2 (2:1 serialization), 4 (4:1 serialization), and 8 (8:1 serialization) for the 8×8 CMP case.



(a)

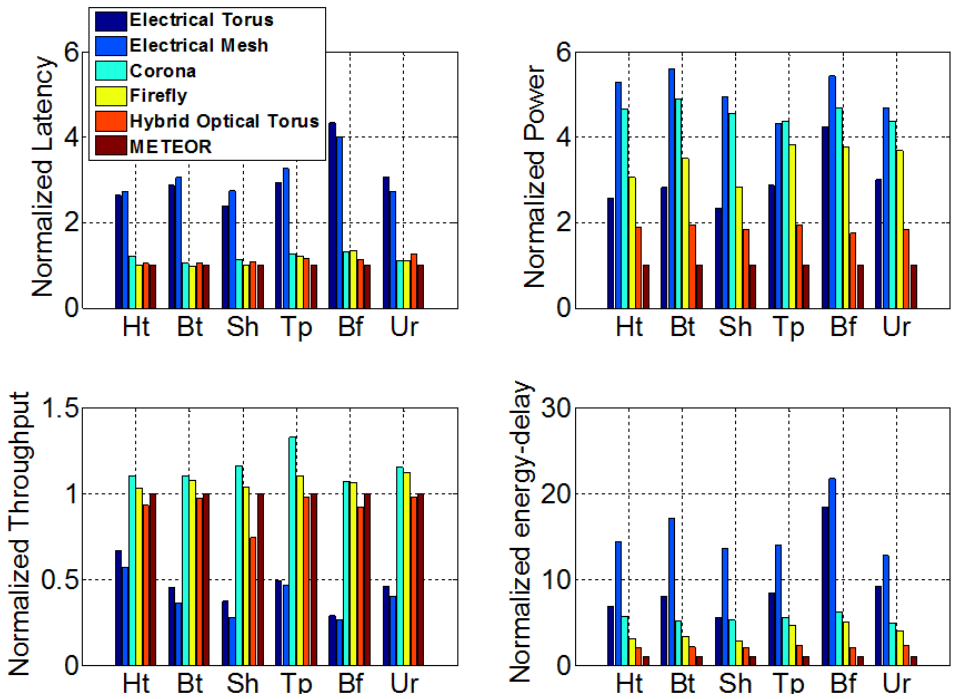


(b)

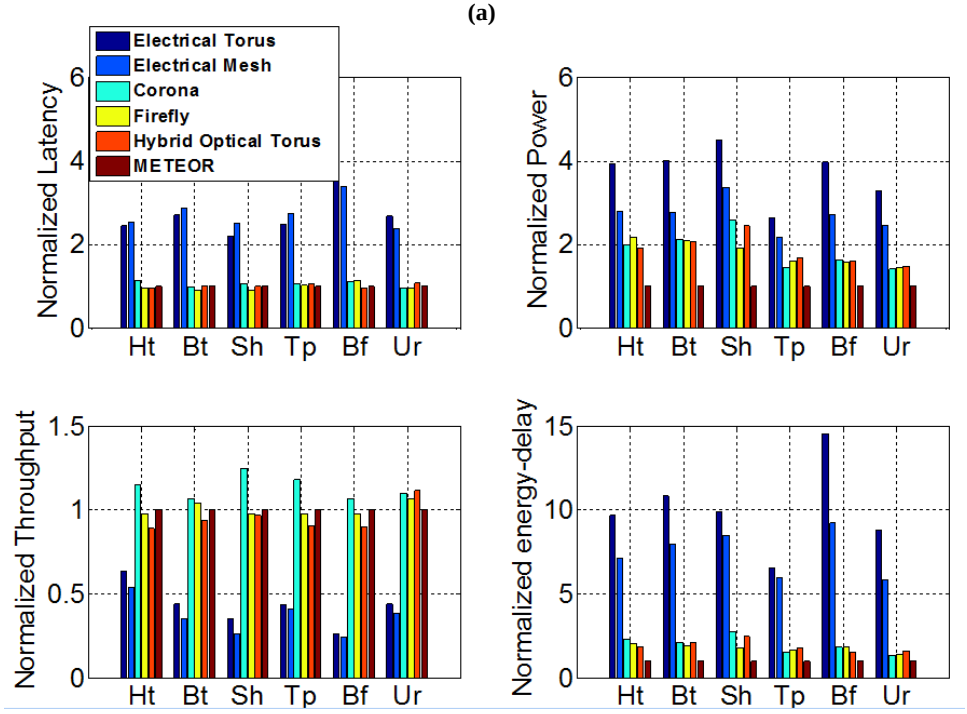


(c)

Figure 25 Impact of varying serialization degree in *METEOR* for 8x8 NoC (a) average latency, (b) throughput (c) power consumption

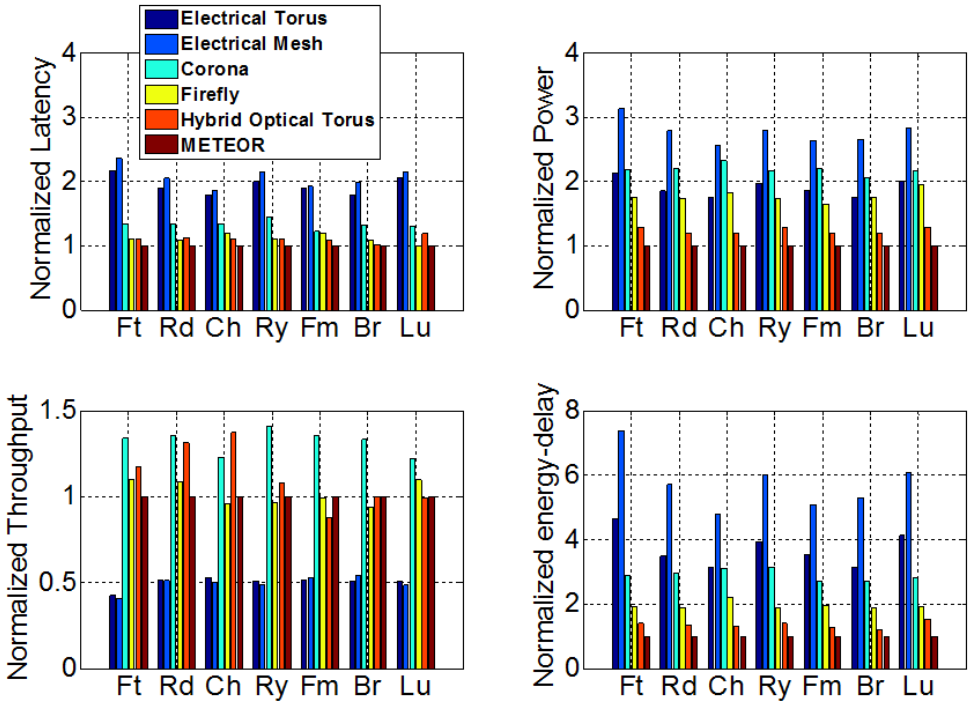


(a)

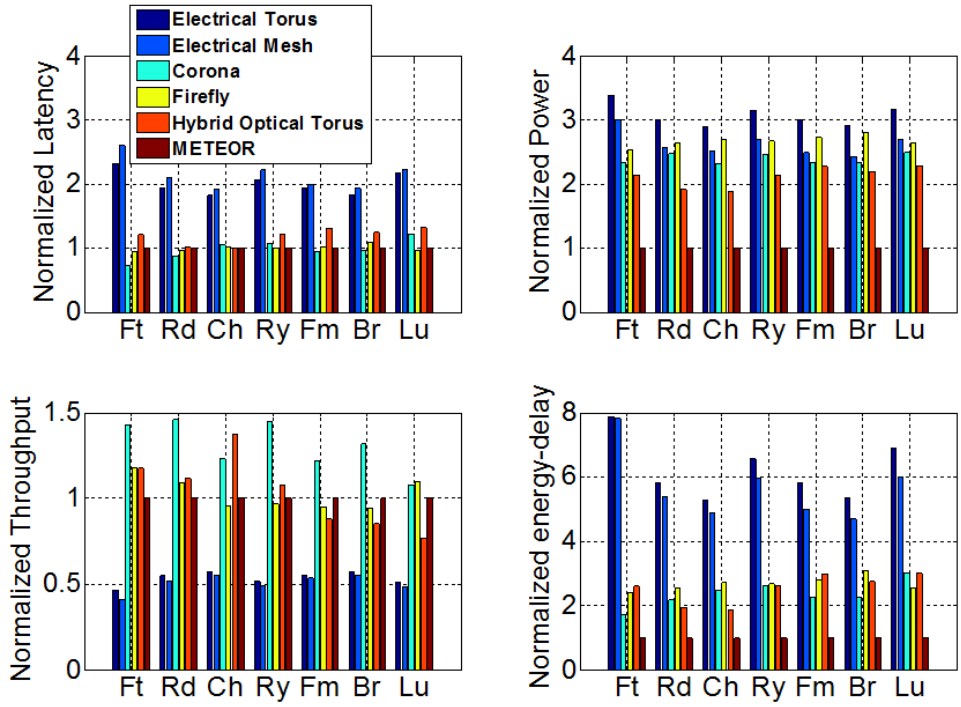


(b)

Figure 26 Normalized latency, throughput, power, and energy-delay product comparison for synthetic benchmarks with (a) 128-bit waveguides and (b) 256-bit waveguides

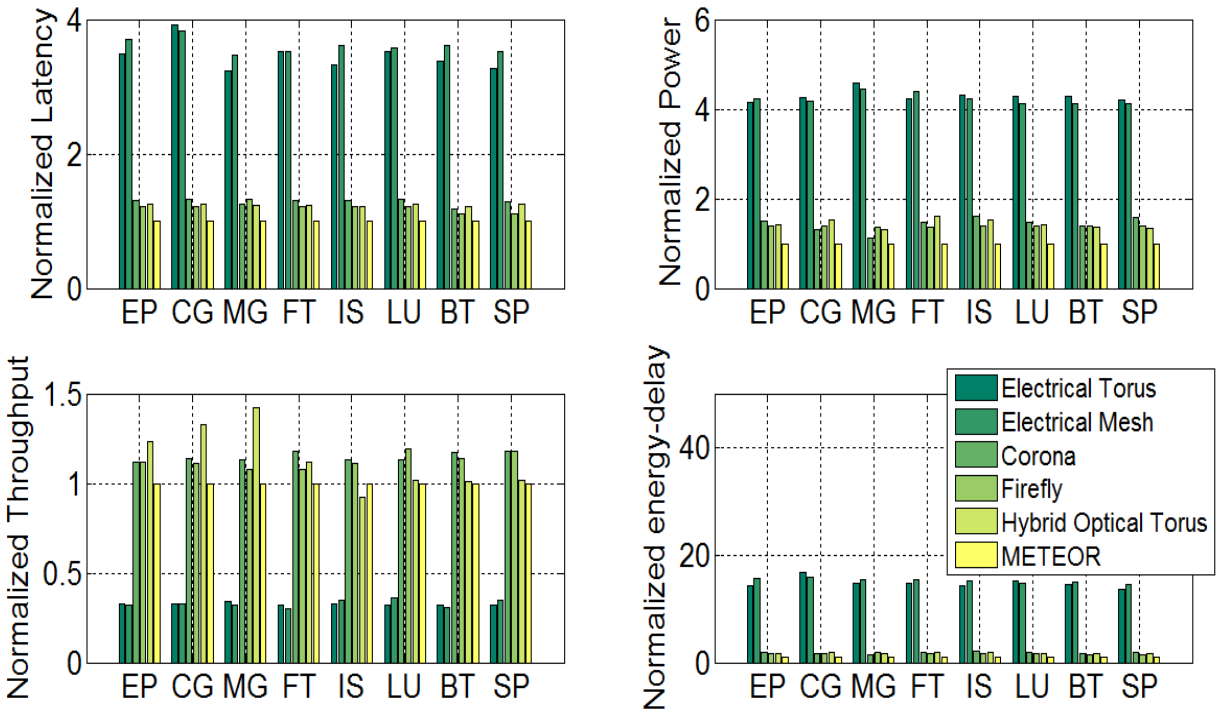


(a)

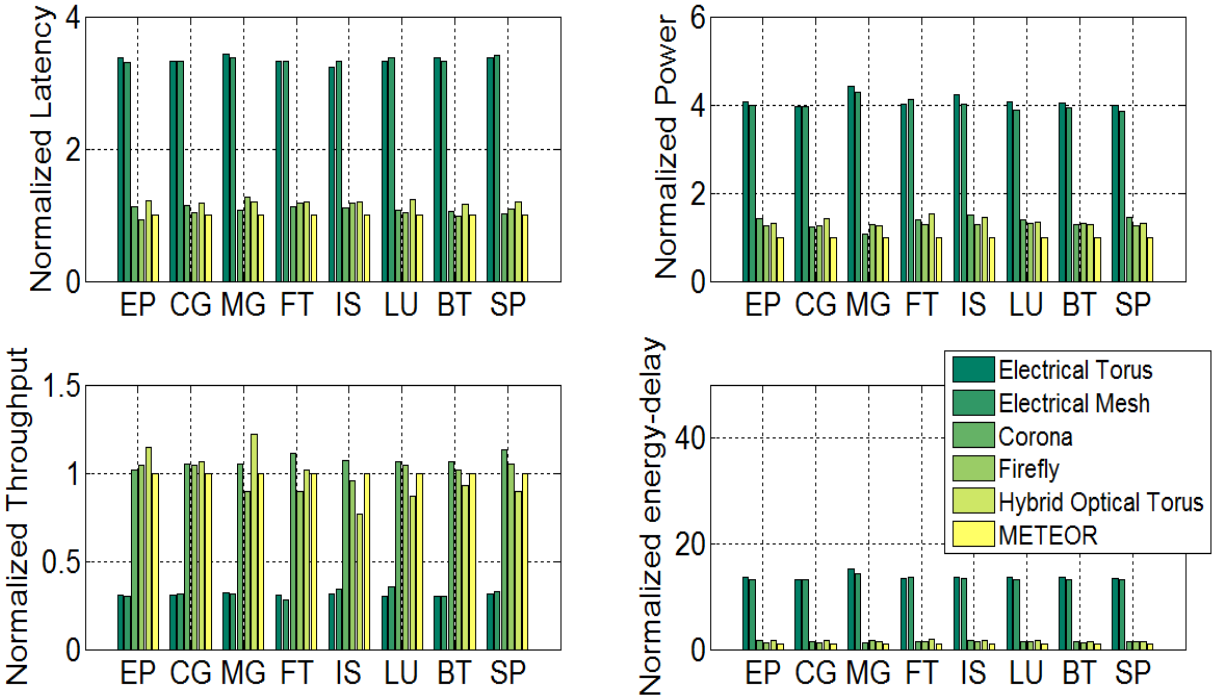


(b)

Figure 27 Normalized latency, throughput, power, and energy-delay product comparison for *SPLASH-2* benchmarks with (a) 128-bit waveguides and (b) 256-bit waveguides

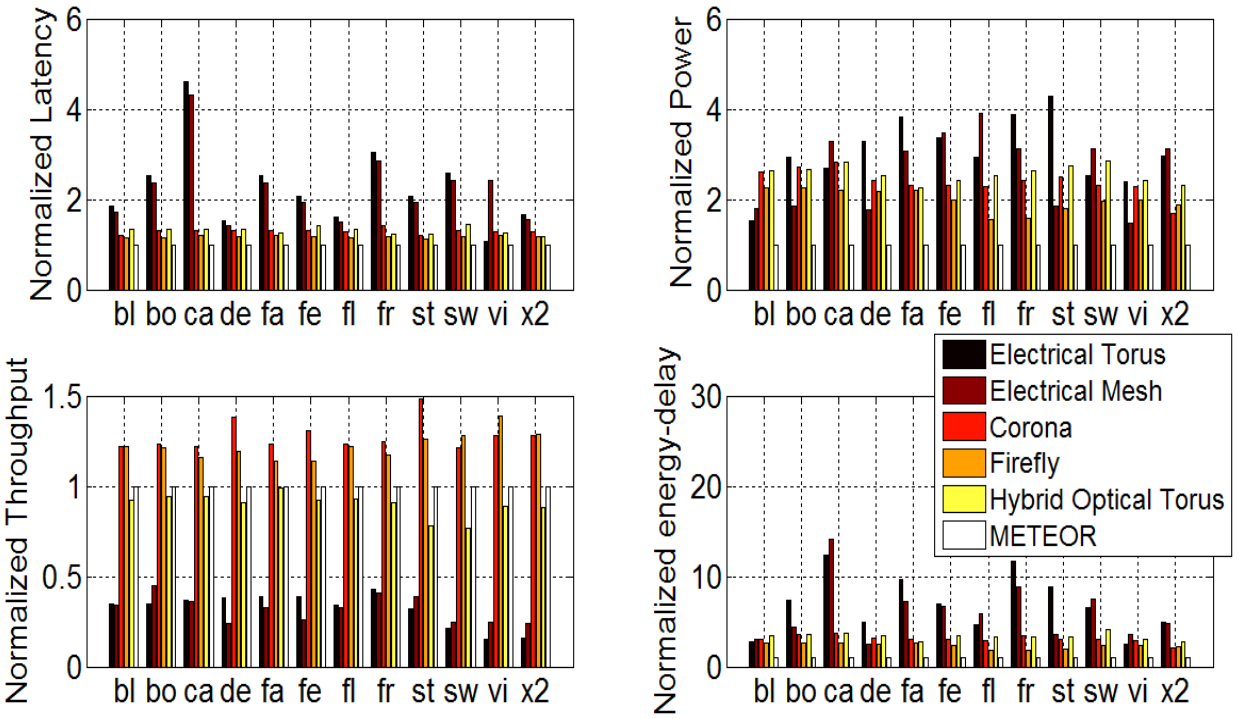


(a)

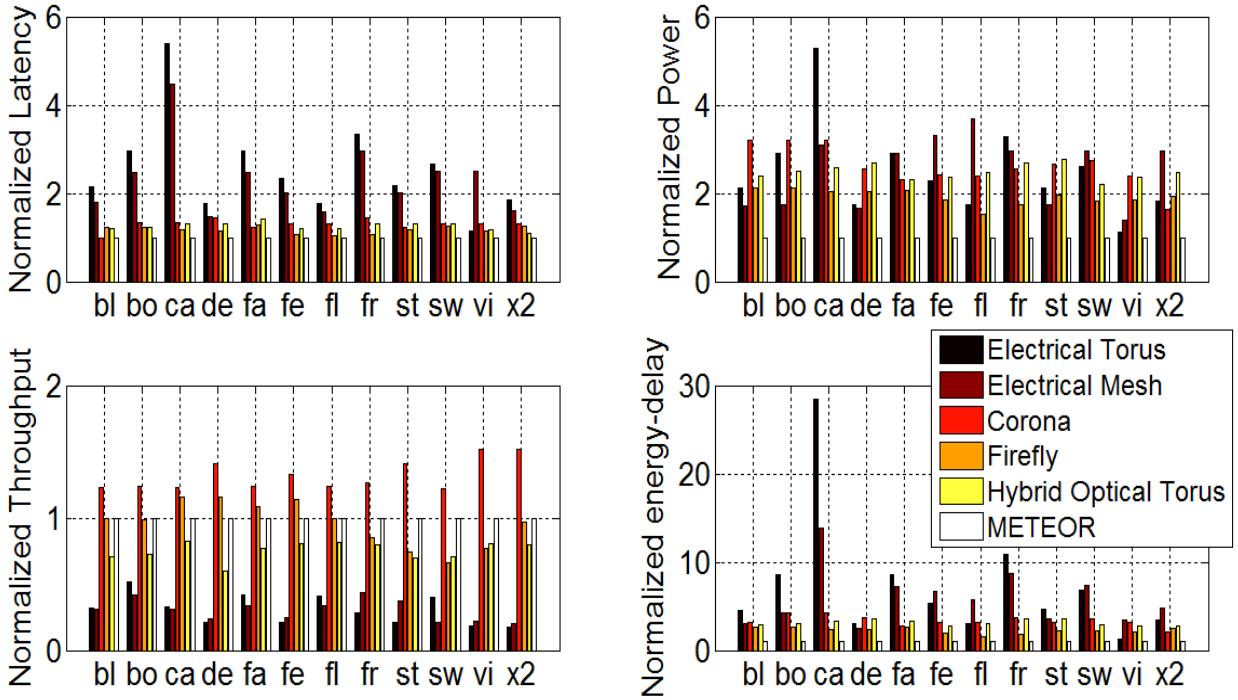


(b)

Figure 28 Normalized latency, throughput, power, and energy-delay product comparison for *NAS* benchmarks (a) 128-bit waveguides and (b) 256-bit waveguides



(a)



(b)

Figure 29 Normalized latency, throughput, power, and energy-delay product comparison for *PARSEC* benchmarks (a) 128-bit waveguides and (b) 256-bit waveguides

From the figure it is clear that serialization has a notable impact on reducing power consumption, due to a reduction in the communication resources and switching activity (even after considering the power consumption overhead of the serializer/de-serializer circuitry). In addition to reducing power, serialization also reduces the number of photonic ring waveguides required, thus reducing the photonic layer complexity, static power consumption and area cost. Serialization however entails a performance overhead as the number of bits transferred in a cycle gets reduced. Figure 25 (a), (b) show the reduction in throughput and an increase in latency as the degree of serialization is increased, for an 8×8 CMP. It is clear that unlike the case where we optimized the PRI size, reducing power with serialization negatively impacts performance. Thus serialization must be used with great care. Serialization degrees of 2 and 4 in particular may provide a reasonable trade-off between power and performance.

3.7.7 COMPARISON WITH OTHER PHOTONIC NOCS

In this section, we compare our proposed *METEOR* architecture with three previously proposed photonic NoCs: *(i)* hybrid photonic torus [102], *(ii)* all-optical Corona crossbar [40], and (iii) hybrid hierarchical Firefly crossbar [37]. We made our best effort to carefully implement every feature of the architectures described in the respective papers for a meaningful comparison. The photonic torus and Corona architectures require photonic conversion even for small local transfers which can be wasteful. The Firefly architecture enables local transfers on the electrical NoC, but only within a concentrated mesh node, while *METEOR* extends the utilization of the electrical NoC further to provide more efficient local transfers within a possibly much larger PRI region, thus reducing load on the photonic layer. Due to the simpler photonic layer architecture, *METEOR* also has *lower power dissipation* compared to other hybrid photonic architectures.

Lower utilization of the electrical NoC also enables much more power efficient and reduced congestion communication in *METEOR* compared to the all-electrical mesh and torus architectures.

Table 3 Micro ring resonator requirement

Relative comparison for photonic resource requirements for 8x8 128 waveguide architecture

Component	***METEOR***	***Corona*** [40]	***Firefly*** [37]	***Optical Mesh*** [102]
Transmission	8192	262144	32768	294912
Reservation	1024	2048	1024	6400
Arbitration	1024	2048	1024	6400
Clock	4	64	16	64
Total	10244	266208	34832	307776

Based on our analysis, we chose a *METEOR* configuration with a PRI size of 16, 32 uplinks, and a serialization degree of 1 to compare against other photonic NoC architectures. The WDM degree for all compared architectures was kept fixed at 64, and results for 128 and 256 bit waveguides were explored for all architectures.

Table 3 shows a relative comparison of photonic microring resonator requirements for an 8×8 core, 128 waveguide architecture. As shown in the table, our proposed *METEOR* architecture requires 3.5 to 25× lower resources compared to previously proposed architectures. The *METEOR* architecture was configured with a PRI size of 16, 32 uplinks, and a serialization degree of 1. The WDM degree for all architectures was fixed at 64, and the results were generated for various synthetic and *SPLASH-2* benchmarks. Figure 26 shows the results for normalized latency, power, throughput, and energy-delay products for synthetic benchmarks: *Hotspot (Ht), Bitwise (Bt), Shuffle (Sh), Transpose (Tp), Butterfly (Bf), and Uniform random (Ur)*. Figure 27 shows the results for the *SPLASH-2* benchmark implementations: *FFT (Ft),*

Radix (Rd), Cholesky (Ch), Radiosity (Ry), Fmm (Fm), Barnes (Br), and Lu (Lu). We also implemented *NAS* [136] and *PARSEC* [137] benchmarks to evaluate our *METEOR* architecture. *NAS* benchmarks are derived from computational fluid dynamics (CFD) applications. The Princeton Application Repository for Shared-Memory Computers (*PARSEC*) benchmark suite is composed of multithreaded programs. *PARSEC* workloads represent next-generation shared-memory programs for CMPs.

Figure 29 shows the results for the *PARSEC* benchmarks (*blackscholes (bl), bodytrack (bo), canneal (ca), dedup (de), facesim (fa), ferret (fe), fluidanimate (fl), freqmine (fr), streamcluster (st), swaptions (sw), vips (vi), ×264 (×2)*). Figure 28 shows the results for the *NAS* benchmarks (*Embarrassingly Parallel (EP), Conjugate Gradient (CG), Multi-Grid (MG), Fourier Transform (FT), Integer Sort (IS), Lower-Upper Gauss (LU), Block Tri-diagonal (BT), Scalar Penta (SP)*). All results are normalized to the results obtained for *METEOR*, and we present results for 128-bit photonic waveguides (Figure 26(a), Figure 27(a), Figure 28(a), Figure 29(a)) and 256-bit photonic wave-guides (Figure 26(b), Figure 27(b), Figure 28(b), Figure 29(b)) for the architectures.

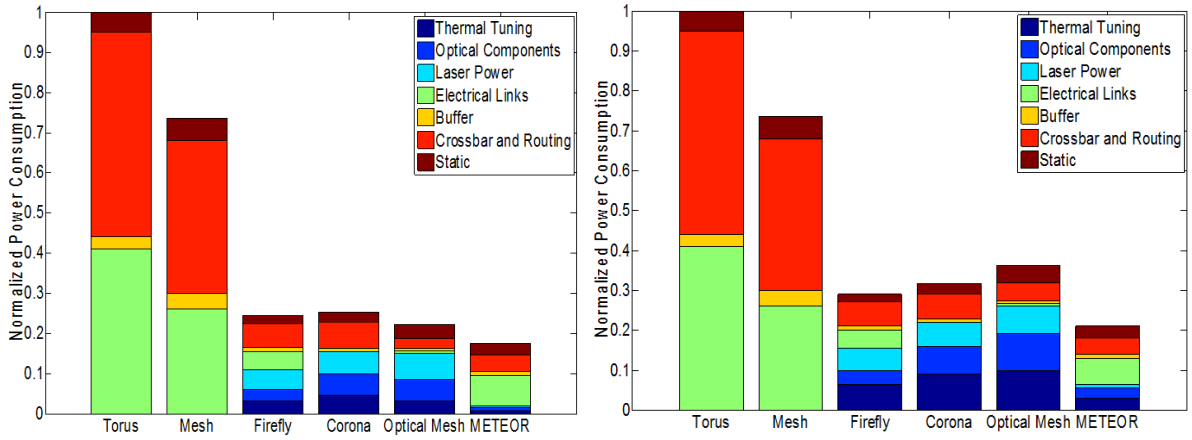


Figure 30 Breakdown of power consumption for (a) 128-bit, (b) 256-bit waveguides.

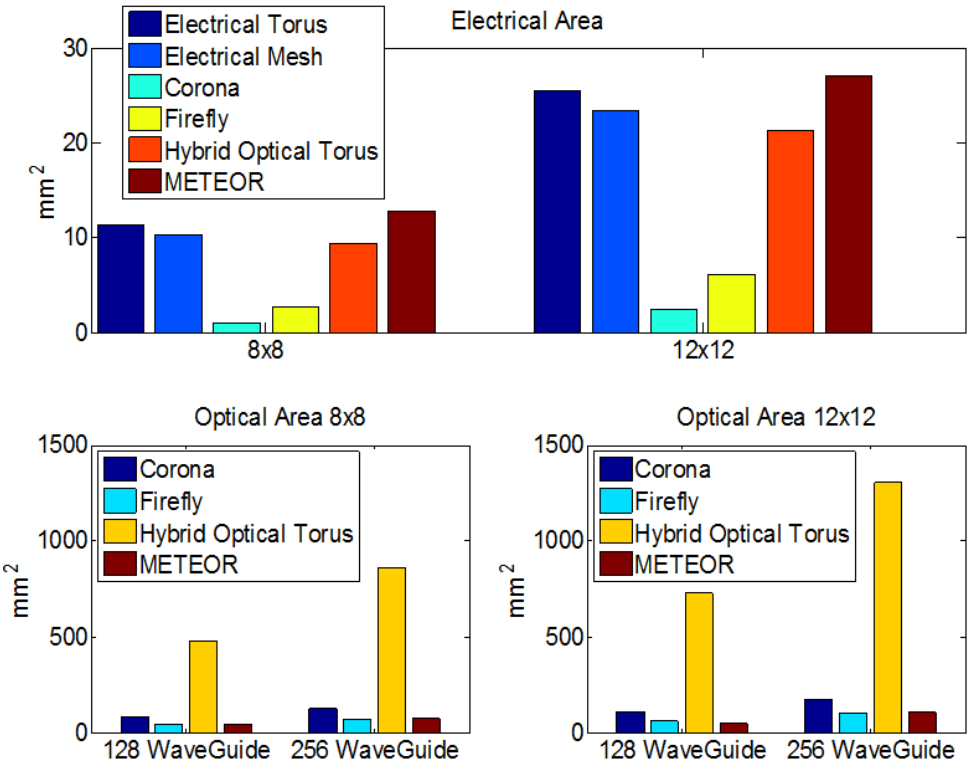


Figure 31 Photonic layer and electrical layer area overhead comparison for various NoC architectures.

Based on Figure 26, Figure 27, Figure 28, Figure 29 *METEOR* configuration with 256 waveguides achieves an average of 89%, 112%, 124% and 33% power reduction compared to previously published hybrid photonic architectures for synthetic, *SPLASH-2*, *PARSEC* and *NAS* benchmarks, respectively. The average latency for *METEOR* is on par with that of other architectures for *SPLASH-2* and synthetic benchmarks. It can be seen that *METEOR* has *slightly lower average latency* for *NAS* and *PARSEC* benchmarks compared to the other architectures because it is able to better balance local transfers over the electrical network with global transfers over the photonic waveguides. For synthetic and *SPLASH-2* benchmarks, *METEOR* achieves comparable throughput with respect to the other architectures while average throughput for

Corona and Firefly was 23% higher than *METEOR* for the *PARSEC* benchmarks. *METEOR* demonstrates an average (256 and 128 waveguides combined) of 73%, 98%, 136% and 58% energy-delay product improvement compared to previously published hybrid photonic architectures for synthetic, *SPLASH-2*, *PARSEC* and *NAS* benchmarks.

Figure 30 shows the breakdown of average power dissipation for the various architectures considered in the comparison study. Results are averaged over the synthetic benchmarks and normalized to the result obtained from the all-electrical torus, for the 128-bit waveguides (Figure 30 (a)) and the 256-bit waveguides (Figure 30 (b)). It can be seen that the *METEOR* architecture has lower laser, photonic component, and thermal tuning power due to its simpler photonic layer architecture. *METEOR* and Firefly have higher electrical link, crossbar, and routing power than other hybrid photonic architectures because both Firefly and *METEOR* support electrical layer data transfers. As the extent of electrical layer communication in *METEOR* is greater than that in Firefly, *METEOR* electrical layer power dissipation is higher than that of Firefly. The static power dissipation in all the NoC architectures is relatively low, due to the communication networks being dominated by dynamic power dissipation.

Figure 31 shows a comparison of the photonic and electrical layer area overheads of the various NoC architectures, for the 8×8 and 12×12 core CMPs. It can be seen that *METEOR* has a higher electrical layer area footprint compared to the other architectures. This area is greater than the electrical mesh area primarily due to the extra router complexity at the gateway interfaces. In the photonic layer, the hybrid photonic torus topology does not scale well with increasing core counts, and has a significant area overhead due to the large number of photonic waveguides and photonic switches. In fact, it requires a multi-photonic layer implementation which will be extremely costly. Firefly and *METEOR* both have lower area overhead than Corona, which uses

significantly higher number of resonators and detector resources. *METEOR* has a comparable optical layer area overhead with Firefly. In summary, *METEOR* provides a notable improvement in average latency, power dissipation, and energy-delay product compared to traditional electrical mesh/torus NoCs and existing hybrid photonic NoCs. Coupled with its low photonic layer complexity, the results motivate considering the *METEOR* hybrid NoC architecture in future CMPs that integrate photonic interconnects.

3.8 RESULT SUMMARY

Future CMP applications with hundreds of cores will require a scalable communication fabric that can enable high performance per watt. It is not clear whether current 2D electrical NoCs can satisfy the performance requirements for future CMP applications with a highly constrained power budget. To address this challenge, in this work we proposed a hybrid photonic NoC *(METEOR)* that combines configurable concentric photonic ring waveguides on a dedicated silicon layer to complement a traditional electrical 2D mesh NoC. Results from our experimental studies indicate that *METEOR* can lead to significant reduction in power consumption and energy-delay product, in addition to improvements in average transfer latency, compared to traditional 2D all-electrical mesh and torus NoCs, as well as the previously proposed hybrid photonic torus, Corona, and Firefly hybrid photonic NoC fabrics. In terms of area overhead, *METEOR* has much lower complexity in the photonic layer, compared to the previously proposed hybrid photonic NoCs. The encouraging results from this work highlight the potential of using photonics on chip and the *METEOR* hybrid photonic NoC architecture to meet the challenges of rising CMP complexity in the future.

4 HYBRID PHOTONIC NOC FOR MULTIPLE USE-CASE APPLICATIONS

Multiple use-case chip multiprocessor (CMP) applications require adaptive on-chip communication fabrics to cope with changing use-case performance needs. In this chapter we propose *UC-PHOTON*, a novel hybrid photonic NoC communication architecture optimized to cope with the variable bandwidth and latency constraints of multiple use-case applications implemented on CMPs. METEOR uses a single nano-photonic ring waveguide based architecture. *UC-PHOTON* extends it to multi-ring architecture providing additional flexibility to configure performance needs. Our detailed experimental results indicate that *UC-PHOTON* can effectively adapt to meet diverse use-case traffic requirements and optimize energy-delay product and power dissipation, with scaling CMP core count and multiple use-case complexity.

4.1 MULTIPLE USE-CASE APPLICATIONS

In recent years, rapid advances in technology scaling and increases in application complexity have given impetus to the design of chip multiprocessors (CMP) with multiple components (processors, memories, peripherals) integrated on a single chip. CMPs can support greater levels of parallelism and have been shown to provide significant improvements in performance-per-watt compared to uni-processor systems-on-chip (SoC) clocked at higher frequencies. Already, numerous CMP designs are commercially available today from several vendors for a wide range of computing systems from Blu-Ray recorders, car navigation systems, digital TVs, gaming consoles, to exa-flop supercomputers. Some examples include the Sony/IBM/Toshiba Cell [5], Fujitsu FR-1000V [148], NEC/ARM MPCore and MP211 [149], and Renesas SH-X3 [150], which have been developed for the consumer electronics domain.

One of the most challenging problems in CMP design today is the design of the on-chip communication fabric that inter-connects the multiple cores on a chip [3]. The on-chip communication fabric is responsible for satisfying strict latency and bandwidth constraints that are relentlessly increasing in stringency as application performance approaches the peta- and exa-flop levels. Unfortunately, on-chip interconnects have not scaled well with process technology. In ultra-deep submicron (UDSM) technology nodes below 65 nm, not only have interconnects become longer, but the signal delay on these long (global) interconnects has been steadily increasing with each successive technology generation, and now far exceeds gate delay. The International Technology Roadmap for Semiconductors (ITRS) acknowledges that delay on global interconnects has now become a major performance bottleneck, and the topmost challenge for the semiconductor industry [4]. In addition to becoming a potential source for performance bottlenecks, interconnects on a chip suffer from reduced reliability due to UDSM effects such as capacitive and inductive crosstalk, and higher dynamic and leakage power dissipation.

To cope with communication demands in emerging CMP designs, there has been a gradual shift away from circuit-switched bus-based communication architectures to packet-switched networks-on-chip (NoCs) [3] [151]. Hierarchical and crossbar-based shared bus architectures lack the scalability to support high bandwidths, and are also more susceptible to intra-die process variations and interference due to crosstalk and external electromagnetic sources.

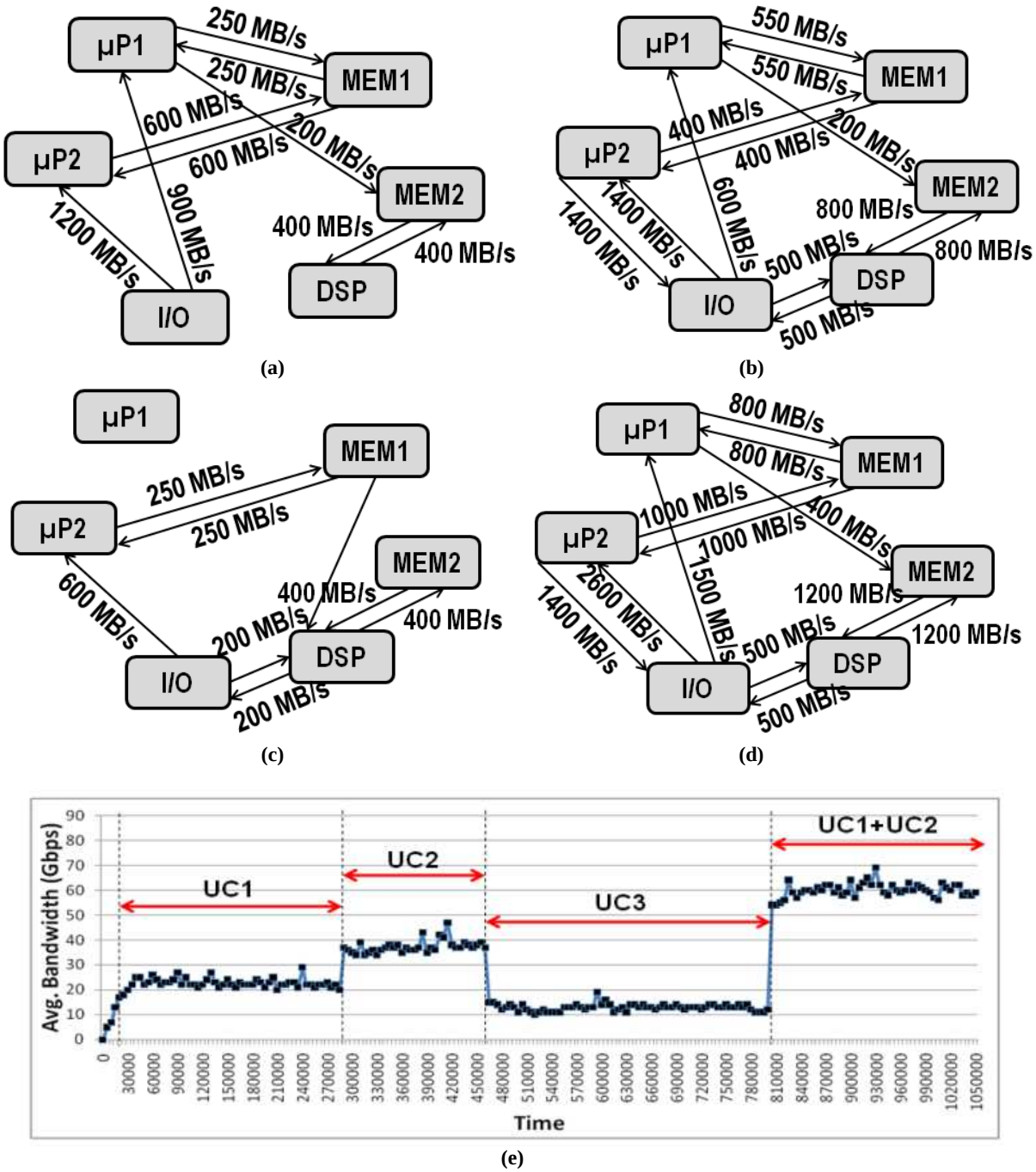


Figure 32 Multiple use case application with use cases (a) UC1, (b) UC2, (c) UC3, (d) UC1+UC2, (e) workloads

NoCs generally offer a more regular structure and a layered design methodology, which leads to better predictability and improved reliability in UDSM technologies. Most importantly, NoCs can support higher bandwidths and are thus being hailed as a promising on-chip

communication fabric for emerging CMP designs. However, in practice, the few existing implementations of NoCs have been found to have several drawbacks stemming from the large number of network interface (NI), router, link, and buffer components in NoCs that lead to a high area overhead and prohibitive power dissipation. For instance, recent NoC prototypes have shown NoCs taking a substantial portion of system power, e.g., ~40% in the MIT RAW chip [152] and ~30% in the Intel 80-core Teraflop chip [152]. Recent studies have also suggested that NoC power dissipation is much higher (by a factor of 10×) than what is needed to meet the performance needs of future CMPs [7] [153]. Thus, radical new solutions are required to overcome the power brick wall facing NoCs in the next few years.

The relatively recent phenomenon of digital convergence [154] puts a further strain on communication architecture design in emerging CMPs. Applications in the digital convergence era have multiple operating modes, called use-cases, that have distinct workload and communication traffic characteristics. For instance, smart cellular phones today can support a combination of several functionalities including making and receiving calls, playback for MP3s, FM radio and video, GPS navigation, PDA support, wireless web browsing, games with 3D graphics, Bluetooth syncing, and so on. Some of the functionalities, such as 3D gaming have a much higher computational and communication workload to render sophisticated graphics and maintain high frame display rates. In contrast, MP3 playback requires simpler audio codec decoding and has a much lower computation and communication performance requirement. With the number of use-cases increasing rapidly in applications today (~tens to hundreds), designers are finding it extremely challenging to create CMP implementations that can support multiple heterogeneous use-cases.

Figure 32 illustrates an example of a network routing application with multiple use-cases. The application is implemented as a CMP with two general purpose microprocessors, a digital signal processor (DSP), two on-chip memories and a network interface (I/O). Figure 32 (a)-(c) show three use cases of the application that have vastly different bandwidth requirements between components. It is also possible for the application to execute multiple use cases simultaneously, as shown in Figure 32 (d), where use cases 1 and 2 execute at the same time (called a compound use case). Figure 32 (e) shows how the average workload traffic bandwidth varies for the use cases during application execution. Typically, switching between use cases takes place when users interact with the application or if there is a change in the environment (e.g., change in wireless signal strength, or battery level). Such a switch in use cases results in the temporal switch of application task and communication graphs. It is highly likely that a communication architecture customized for a single use case may not meet performance requirements for another use case. Thus there is a need to enhance traditional on-chip communication fabrics for multiple use-case applications.

In this chapter, we propose using a novel hybrid photonic NoC communication architecture called *UC-PHOTON* to cope with emerging multiple use-case applications and maximize performance-per-watt. *UC-PHOTON* is comprised of one or more photonic ring paths coupled to a traditional 2D electrical mesh NoC architecture. The photonic paths offload global communication from the electrical network, improving packet latency and reducing communication power dissipation. In addition, *UC-PHOTON* supports dynamic reconfiguration of the electrical and photonic networks. This enables runtime adaptation to changing traffic patterns, which allows network resources to be optimized for even lower power dissipation. Experimental results on several multiple use-case CMP benchmark implementations indicate that

UC-PHOTON can scale with increasing use-case count and core count to save orders of magnitude power, reduce energy-delay product, and improve performance compared to traditional electrical NoC architectures.

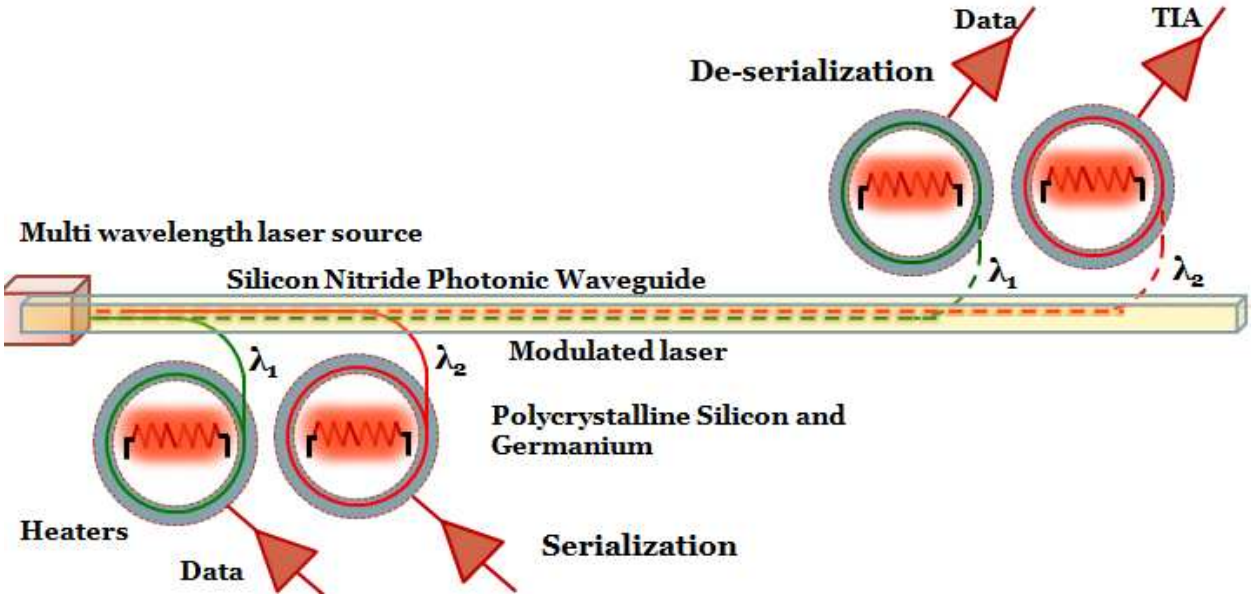


Figure 33 Building blocks of on-chip photonic interconnects

4.2 ON CHIP PHOTONIC ARCHITECTURE OVERVIEW

Recently, photonic interconnects have been proposed as a solution to overcome the on-chip communication power bottleneck [29]. Photonic interconnect technology is of interest because it has been shown to be much more energy efficient compared to copper (Cu) interconnects especially at high speeds and long distances [6] [89] [92]. The ability of photonic waveguides to carry many information channels simultaneously increases interconnect bandwidth density significantly, eliminating the need for a large number of wires to achieve adequate bandwidth. Photonic interconnects are becoming standard in data centers, and chip-to-chip photonic links

have been demonstrated [155]. This trend will naturally bring photonic interconnects into the on-chip stack, particularly as a means to enable high bandwidth and low power data communication between hundreds of cores in future CMPs. Recent advances in the field of nanoscale silicon photonics have enabled highly integrated photonic interconnect-based components in CMOS-based ICs [11] [13] [14]. In fact, photonic elements have now become available as library cells in standard CMOS processes.

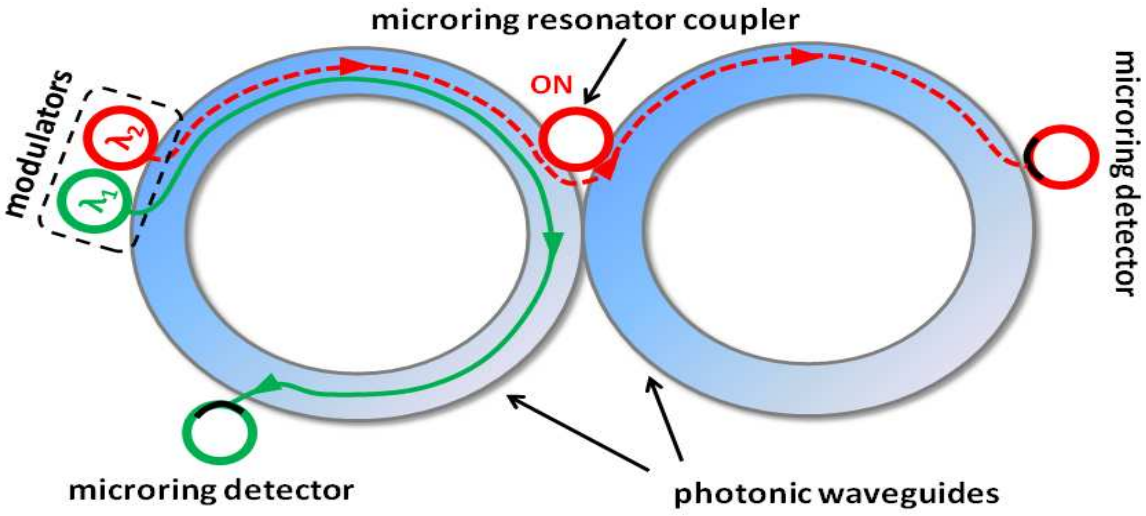


Figure 34 Microring resonator coupling for multi ring waveguides

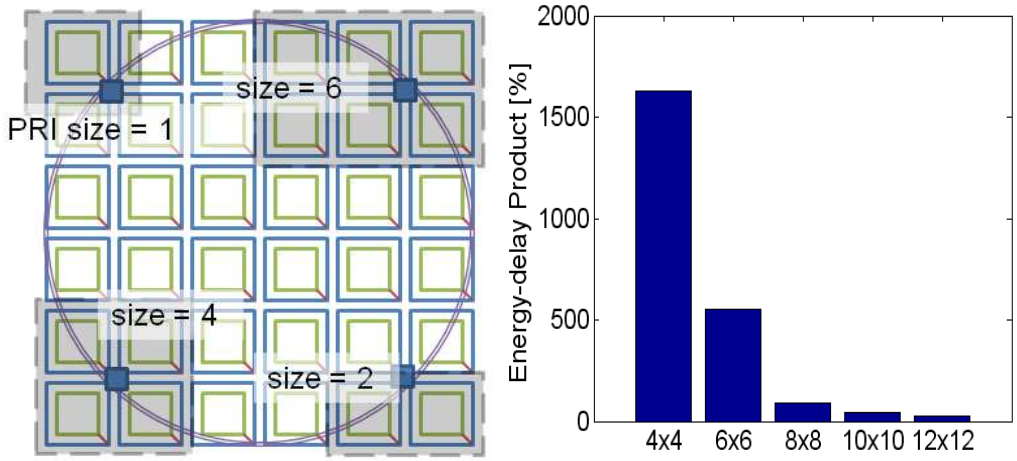
Figure 33 shows a high level overview of the key on-chip photonic transmission components as considered in this work: a multi-wavelength laser light source, modulator, photonic waveguide, and photodetector. Multiple wavelengths of light are coupled on to the chip using photonic fibers from off-chip (or on-stack) multi-wavelength mode-locked lasers [24]. Several simultaneous data transfers over multiple wavelengths are possible in photonic interconnects, and such wave division multiplexing (WDM) is critical for ensuring high

bandwidth transfers [156]. For a WDM scheme with λ wavelengths, allocation of wavelengths to traffic streams is done using ‘multiplexing by core’, with each of the n interfacing cores having exclusive access to λ/n wavelengths. This limits the number of transmitters, but provides substantial power savings. The photonic waveguide is made of CMOS-compatible silicon oxide, which has been shown to carry light with low losses (on the order of 2–3 dB/cm) and can be curved with bend radii on the order of 10 µm [144].

Wavelength-selective nanophotonic silicon modulators made out of microring resonator structures [26] are used to convert electrical signals into light at the source, for transmission. Microring resonators are also used at the destination as filter structures to “drop” the corresponding wavelength from the waveguide into a local photodetector device [27] that converts the light signal back to an electrical signal. Trans-impedance amplifier (TIA) circuits are used to amplify analog electrical signals at the receiver to digital voltage levels. Future CMPs with hundreds of cores will require multiple photonic waveguides to meet performance requirements. In such systems, microring resonators can be used as couplers to couple light between multiple waveguides. Figure 34 shows an example of this coupling, with light of wavelength λ_2 being coupled from the first to the second photonic waveguide. The coupling is enabled by injecting charge into the microring resonator coupler to change its index of refraction so that only a resonant wavelength λ_2 is coupled through it, while other non-resonant wavelengths (such as λ_1) remain unaffected. Ring filters and modulators must also be thermally tuned to maintain their resonance under on-die temperature variations [36]. We assume a single heater element per microring resonator structure in this work for this purpose.

An important consideration for WDM enabled photonic interconnects is the optical loss in its components. Optical loss impacts system design as it sets the required optical laser power and

correspondingly the electrical laser power (at a roughly *30%* conversion efficiency). In addition to the waveguide, optical losses exist for couplers, modulator/filter resonators, waveguide crossings, photodetectors, and also due to non-linearity.



(a) **(b)**

Figure 35 (a) 6x6 hybrid NoC with photonic ring and various sizes of photonic regions of influence (PRI), (b) % improvement in energy-delay product for hybrid NoC with photonic ring compared to conventional 2D mesh NoC, with scaling CMP complexity.

4.3 *UC-PHOTON* OVERVIEW

4.3.1 BACKGROUND

Our proposed *UC-PHOTON* communication architecture utilizes 3D integration with separate planes for logic and silicon photonics as presented in Figure 6. The basic architecture comprises of an electrical NoC interfaced to a silicon photonics layer that has photonic waveguide-based interconnect paths. In general, photonic waveguides and components for complex topologies such as mesh, torii, and fat trees can be prohibitively expensive in terms of fabrication cost and area overhead. Consequently, in our previous work [44], we proposed a low overhead hybrid photonic NoC architecture with a parallel ring-based photonic waveguide

interfaced to a traditional 2D electrical mesh NoC. Gateway interface routers provided the connectivity between the electrical layer and the modulators and photodetectors in the photonic layer. The photonic ring was shown to provide a faster and more energy-efficient path for on-chip global communication compared to traditional electrical NoCs.

To improve scalability, a photonic region of influence (PRI) was also defined in [44], which refers to the number of cores around the gateway interface that can utilize the photonic path for communication. Figure 35 (a) shows a 6×6 CMP with varying PRI sizes at the four gateway interfaces. If a router falls under a PRI, it is modified to additionally consider the photonic path for global communication for incoming flits. Note that while the sizes for the PRIs are shown as different at each gateway interface in the figure, this is for illustration purposes only, and in practice we assume a fixed PRI size for all gateway interfaces. For smaller sized systems (e.g., 4×4 CMPs), limiting the number of cores interfacing with each gateway interface to one (i.e., PRI size = 1) may be sufficient to offload a majority of the global communication from the electrical network. For more complex systems (e.g., 8×8 CMPs) a larger PRI size may be more appropriate. However, we have found that as the system size increases (e.g., 10×10, 12×12 cores etc), increasing the PRI size provides rapidly diminishing returns.

Figure 35 (b) shows the % improvement in the energy-delay product of the hybrid photonic-ring NoC compared to a conventional electrical mesh NoC, with scaling CMP complexity. For each CMP configuration, results were averaged over several *SPLASH-2* benchmark [135] implementations, and optimal PRI sizes (to get the lowest energy-delay product) were used. It is clear from the figure that the benefits of using the photonic ring become insignificant for large CMP sizes. This is primarily because the photonic ring is under-utilized due to its limited coverage area on chip, even though the global communication requirements are

higher. Thus, more scalable hybrid communication fabrics are needed to cope with the more stringent latency and power constraints of future CMPs.

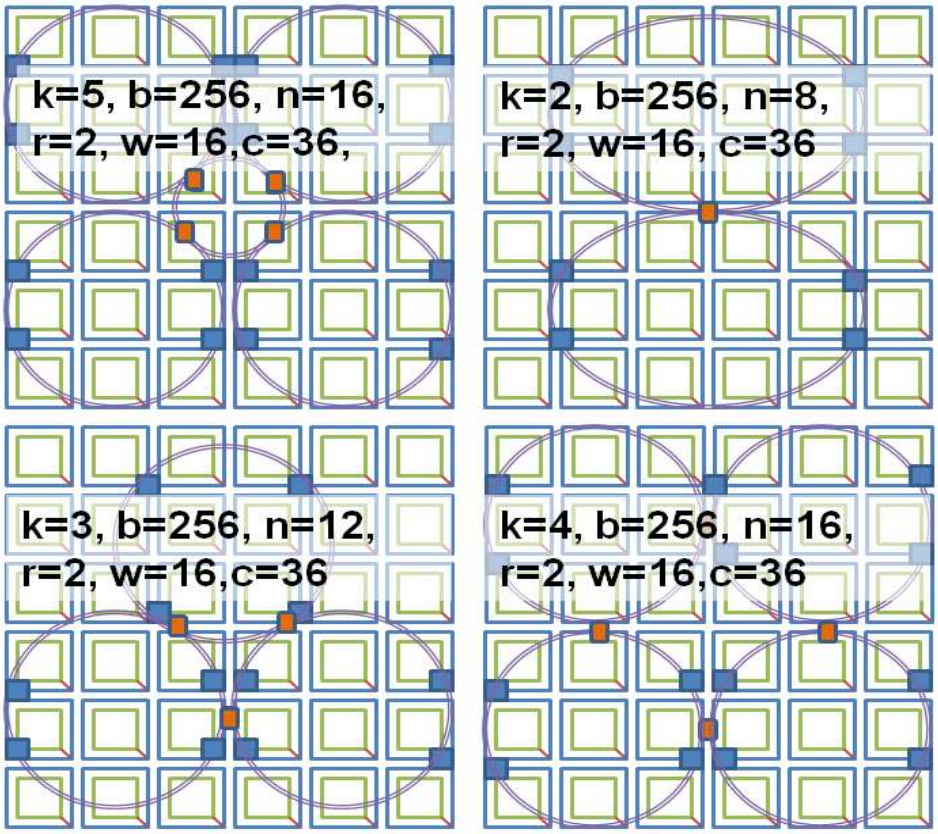


Figure 36 Four configurations of the proposed hybrid photonic NoC architecture for a 6x6 CMP

4.3.2 TOPOLOGY

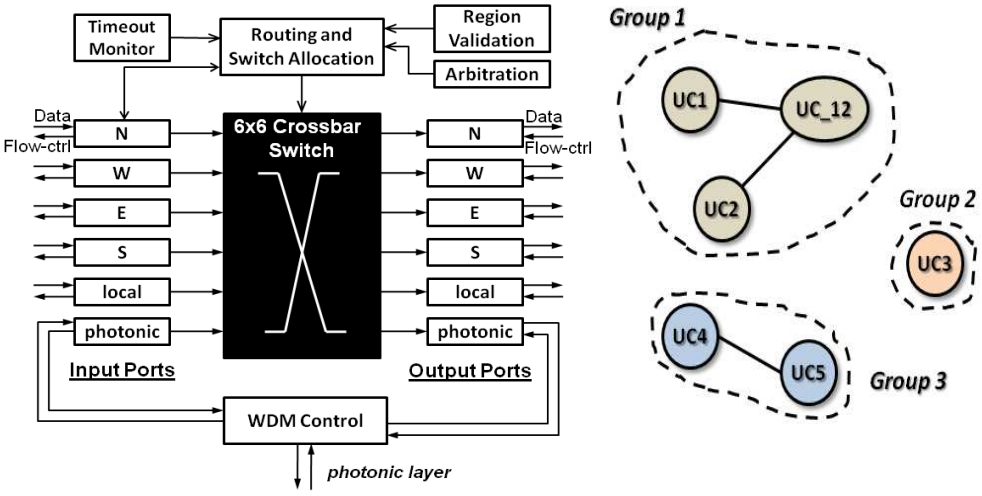
To overcome the scalability limitations discussed above, we propose a more general hybrid photonic NoC architecture called *UC-PHOTON* that employs multiple photonic ring paths. The topology can be characterized by a 6-tuple <k,b,n,r,w,c> where k is number of photonic rings, b is the bitwidth of the photonic waveguides, n is number of uplinks/downlinks (i.e., gateway interfaces), r is the PRI size, w is the number of WDM channels, and c is the number of cores in the CMP. For the purpose of this work, we analyzed various 6-tuple spaces to better understand

the scalability of various configurations. Figure 36 shows four configurations of the proposed architecture with varying numbers of photonic ring paths (k = 2,3,4,5; with n = 8,12,16,16 gateway interfaces respectively) that can better span an IC chip. The configurations shown are for a fixed sized CMP (c = 36), have a bitwidth b = 256, PRI size r = 2, and w = 16 WDM channels. The building blocks of the photonic communication, such as the modulators, waveguides, photo-detector receivers, and microring resonator switches are the same. The electrical mesh NoC requires modifications in the router architecture to incorporate additional traffic to/from the photonic layer. This is discussed in detail next.

4.3.3 ROUTING AND FLOW CONTROL

The *UC-PHOTON* communication architecture supports wormhole switching, XY dimension order routing for routing flits in the electrical NoC, and a modified PRI-aware XY routing scheme for selective data transmission through the photonic links. Communicating cores lying within the same photonic region of influence communicate using the electrical NoC (intra-PRI transfers). Cores that need to communicate and reside in different photonic regions of influence communicate using the photonic paths, if the size of the data is above a certain user-defined threshold (inter-PRI transfers). In this way large data messages can be offloaded from the electrical NoC and sent over a faster and more energy efficient photonic path. Transfers between cores lying outside photonic regions of influence occur normally via the electrical network. Network interfaces (NIs) ensure that header flits contain coordinates of the source and destination of the packet being injected into the NoC, as well as a flag indicating that the message size is large enough to potentially traverse a photonic path (for inter-PRI transfers). All routers have ‘region validation’ units that use XY routing for intra-PRI transfers or for transfers

to cores not residing in any PRIs via the electrical NoC. Otherwise if an inter-PRI transfer is detected by the 'region validation' unit at the source router, the packets are re-routed to the gateway interface of the local PRI using XY routing, traverse the photonic link to the destination gateway interface, and then are routed to the destination router using XY routing. The data traversing the photonic path is not buffered in the photonic layer, and the photonic transfer can therefore be considered to be a form of photonic circuit switching. Flow control for these transfers is implemented using the switch-to-switch ACK/NACK scheme in both the electrical and photonic links, to ensure that the destination is able to accept the transmitted data. Flits traversing a photonic link receive ACK/NACK information from the destination via the electrical NoC, just like with electrical links.



(a) **(b)**

Figure 37 (a) Gateway interface electrical router architecture, (b) use-case critical transition graph (CTG)

There are two types of routers used in *UC-PHOTON*: *(i)* regular electrical mesh routers that have *5 I/O ports (N, S, E, W, local core)* with the exception of the boundary routers that have fewer ports, and *(ii)* gateway interface routers that have six I/O ports (*N, S, E, W, local core,*

photonic link) and are responsible for sending/receiving flits to/from photonic interconnects in the photonic layer. Both types of routers have an input and output queued crossbar with a 4-flit buffer on each input/output port. Figure 37 (a) shows a high level block diagram of the gateway interface router. Note the additional ports for interfacing with the photonic rings. These ports are connected to a 'WDM control' module that controls wavelength assignment to different traffic flows, to enable WDM for high bandwidth photonic communication. If multiple requests contend for access to the photonic waveguide at a gateway interface, then the request with the furthest distance to the destination is given preference (other schemes can be used here as well).

While XY routing has been proven to be deadlock-free for mesh-like regular NoCs (as no channel dependency cycles can be formed between dimensions), the modifications made to this routing scheme in our architecture to accommodate photonic transfers may end up creating deadlock conditions. We extensively studied deadlocks in the proposed architecture when packets traverse the photonic ring paths. To overcome a potential deadlock, we arrived at using low overhead timeout flits sporadically interleaved with the flits for the long data messages traversing the photonic paths. This is a form of regressive deadlock recovery [125]. If a timeout flit reaches a router where flits are blocked, a 'timeout monitor' module in the router can detect a timeout event and recognize potential cases where flits are blocked due to deadlock, and drop the blocked flits, while sending a NACK signal in the reverse direction to indicate the flits being dropped. This allows the system to unblock and recover from potential deadlock. While the method has the overhead of the additional flits in long messages intended for photonic links and monitoring module in the routers, this is still simpler than other potential deadlock resolution alternatives such as keeping a reserved deadlock free channel and draining deadlocked packets through this channel until the deadlock condition clears.

4.3.4 DYNAMIC CONFIGURATION

A key feature of *UC-PHOTON* is the support for dynamic reconfiguration for low power operation. Since different use-cases have different traffic flows, and bandwidth/latency requirements, runtime reconfiguration strategies can be employed at the transition point between use-case executions to save power. There is always a timing overhead incurred when transitioning between use-cases, primarily to load the new use-case data and code, distribute control signals across the chip, and gracefully shut down the current use-case. Some use cases are critical and need to be loaded and run quickly. For other use cases, the transitioning time can be much longer, from hundreds of micro-seconds to several milli-seconds. In this time we can implement strategies to reduce power dissipation. Obviously, the runtime reconfiguration strategies possible during critical use-case switching are much more restricted than for regular use-cases switching. We define a critical transition graph as an undirected graph *CTG(V,E)* where each vertex $vi \in V$ is a use case, and an undirected edge (v_i, v_j) represents the need for fast switching between use cases v_i and v_j. Figure 37 (b) illustrates a CTG for a multiple use-case application, where for instance use cases UC4 and UC5 need to implement fast switching.

For critical use-case transitions, low overhead runtime reconfiguration schemes are desirable to save power. *UC-PHOTON* employs low-overhead runtime optimization techniques *(i)-(iii)* described below for critical use-case transitions, and all the techniques *(i)-(v)* for non-critical use-cases transitions:

4.3.4.1 DVS/DFS

Dynamic supply voltage and clock frequency scaling (DVS/DFS) is one of the most widely used runtime optimization techniques to reduce power dissipation. In our approach, NoC link

and router frequencies are dynamically adapted to meet performance requirements while consuming the minimum power. Since use-case performance requirements are known in advance, the available slack can be precisely utilized to achieve maximum power dissipation. An almost quadratic reduction in dynamic power dissipation can be achieved using this approach. We use a conservative model for voltage scaling, where we assume that the square of the voltage scales linearly with the frequency [157].

4.3.4.2 CLOCK GATING

For some use cases, not every link or router of the NoC needs to be active to implement all communication flows. In such a scenario, to reduce dynamic power dissipation, clock gating can be employed. Clock gating is the most effective solution for optimizing the dynamic power, and is supported by most commercial synthesis and optimization tools. The main idea behind clock gating is to shut-down a circuit's blocks that are not performing useful computations during some particular clock cycles. In our approach, clock gating is used to shut down links and routers that do not need to be accessed for a use-case.

4.3.4.3 DYNAMIC WDM

Wavelength division multiplexing allows several photonic signals to be transmitted simultaneously in a single photonic waveguide using different wavelengths which do not interfere with each other. WDM can thus significantly improve photonic interconnect bandwidth density over electrical interconnects. We assume that each of the photonic waveguides has λ available wavelengths for WDM, thus creating a λ-way WDM photonic path. The value of λ has significant implications for performance, cost and power since using a larger number of

wavelengths improves bandwidth but requires additional modulators and receivers, which increases area, cost and power overhead in the photonic layer. In our work, we utilize a practically achievable conservative λ value of 16. Since the dissipated power in the modulators and receivers is typically a linear function of the number of WDM channels employed, reducing the number of WDM channels can save power. Our hybrid photonic NoC supports rapidly varying the number of WDM channels during use-case transitions, by shutting off channels (modulators/receiver pairs) when data bandwidth requirements are low to save power, and enabling the channels when bandwidth requirements become high, to maintain performance goals.

4.3.4.4 RUNTIME PRI RECONFIGURATION

The size of the region of photonic influence impacts the photonic path utilization. Small region sizes promote more transfers via the electrical NoC, while large region sizes increase the traffic flows eligible for transfer via the photonic rings. However, increasing the PRI size beyond a certain point can be counter-productive, increasing latency and power dissipation, as shown in our previous work [44]. Since application traffic characteristics can change across use-cases, a single PRI size may not be adequate in optimally distributing traffic flows between the electrical and photonic paths to minimize power dissipation. Thus *UC-PHOTON* supports dynamically varying the PRI size at runtime during non-critical use-cases transitions to track changing application traffic requirements, and achieve low power operation. The reconfiguration step involves updating region boundary coordinates in tables in the ‘region validation’ units of the NoC routers, which can take several hundreds of cycles, and hence is only applied during non-critical use-case transitions.

4.3.4.5 ADAPTIVE TDMA SLOT ALLOCATION

The TDMA slot allocation in a router for different traffic flows controls the bandwidth and also the average latency of packets for the flows in a NoC. Since different use-cases have different bandwidth and latency requirements, changing the TDMA slot allocation during use-case transition is a way to adapt to the new use-case. The goal is to give more slots to critical traffic flows in a use-case with more stringent constraints. Since the TDMA slot allocation information is assumed to be stored in a separate memory, it takes several hundreds of cycles to update the TDMA slot allocations in all the NoC routers. Thus, this technique is only applied during non-critical use-case transitions.

Table 4 Multi use-case application characteristics

Application	Cores	Use cases	Critical Trans
SPL2xA	6×6	5	16
SPL2xB	6×6	10	32
PNET1	8×8	15	44
PNET2	8×8	20	54
PNET3	10×10	30	98

4.4 EXPERIMENTS

4.4.1 EXPERIMENTAL SETUP

Photonic waveguides provide faster signal propagation compared to electrical interconnects because they do not suffer from RLC impedances. But in order to exploit the propagation speed advantage of photonic interconnects, electrical signals must be converted into light and then back into an electrical signal. This process requires a performance and power overhead that must be taken into account for an accurate analysis. To explore the impact of using *UC-PHOTON* in CMPs, we modeled it by extensively modifying our in-house cycle accurate SystemC-based NoC

simulator [134] [140] [158] [159]. Five multiple use-case benchmark applications were selected and implemented on multiple cores in the simulator model. The goal was to explore applications with a wide spectrum of core count and use-case complexity. Table 4 shows the characteristics of these applications, with core complexity varying from 36 to 100 cores and use-case complexity varying from 5 to 30. SPL2xA and SPL2xB are multiple use-case applications derived from combining *SPLASH-2* benchmarks (*Cholesky, FFT, Fmm, Lu, Ocean, Radix*). PNET1, PNET2, and PNET3 are multiple use-case networking applications used for packet processing, and forwarding [118]. We targeted a 32 nm process technology, and assume a 400 mm2 CMP die area. A high level floorplanner [134] is used to determine core placement and link lengths.

Table 5 Delay of PHOTON components at 32nm

Component	32 nm
Modulator driver (ps)	9.5
Modulator (ps)	3.1
Waveguide (ps/mm)	15.4
Photo Detector (ps)	0.22
Receiver (ps)	4.0

The operating frequency of the photonic rings was estimated by calculating the time needed for the light to travel from any node to the farthest node, so that data can be transmitted to all nodes in one cycle. Through geometric calculations for the ring, using delay values from Table 5, and incorporating latching delays (using ITRS data [4]) we obtained a maximum operating frequency of greater than 3 GHz for the different sizes of CMPs we considered. Thus the photonic rings (and the communication network) were safely clocked at 2.3 GHz. Delay estimates for the various photonic interconnect-centric components were obtained from [89] and device fabrication results (e.g., [140]), and are shown in Table 5, for the *32 nm* node. The delay

of an optimally repeated and sized electrical (Cu) wire at 32 nm was assumed to be *42 ps/mm* [29]. Delays for other electrical NoC components (routers, NI, buffers) were obtained from post-synthesis gate-level models after layout.

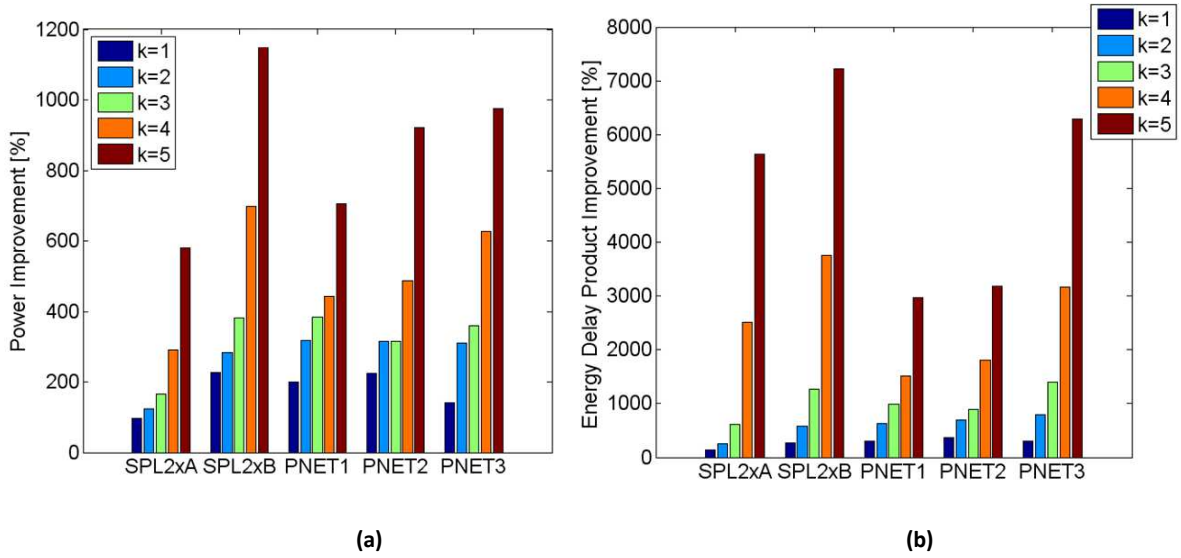


Figure 38 % Improvement for non-reconfigurable *UC-PHOTON* vs. 2D electrical mesh NoC, (a) power, (b) energy-delay product

4.4.2 RESULTS

The power dissipated in *UC-PHOTON* can be categorized into two components: electrical network power and photonic ring network power. The static and dynamic power dissipation of electrical routers and links in this work is based on results obtained from Orion 2.0 [141] incorporated into our simulator. For calculating power dissipation of the modulator driver and TIA power we used ITRS device projections [4] and standard circuit procedures. The energy consumption of each transmitter and receiver is *20 fJ/bit* (dynamic) and *5 fJ/bit* (static), as derived from [36]. A thermal tuning energy of *16 fJ/bit* is considered for each heater element. In addition, an electrical laser power of 3.3 W (with 30% laser efficiency) is also considered in the power calculations. The laser power value accounts for the per component optical losses for the

coupler/splitter (1.2 dB), non-linearity (1 dB at 30mW), waveguide (3 dB/cm), waveguide crossings (0.05 dB), ring modulator (1 dB), receiver filter (1.5 dB) and photodetector (0.1 dB).

Our first set of experiments compares the *UC-PHOTON* architecture with a traditional electrical 2D mesh NoC. Figure 38 (a)-(b) show the improvement in average power and energy-delay product for our proposed *UC-PHOTON* architecture without enabling any runtime optimizations.

The results are shown for the five different applications (from Table 4) implemented on the five different configurations of our proposed architecture (k(n)=1(4), 2(8), 3(12), 4(16), 5(16); b=256; r=optimal; w=16) and the 2D electrical mesh NoC. All implementations represented by bars in Figure 38 satisfy the bandwidth and latency constraints of all use-cases in the respective applications. It can be seen that even without enabling any runtime reconfiguration, *UC-PHOTON* achieves improvements from 2.1× to 12.6× in power saving and a reduction from 3.1× to 73.2× in energy-delay product over a traditional 2D electrical mesh NoC. These significant improvements are due to the offloading of multi-hop communication from the electrical mesh path to the more energy efficient and lower latency photonic path. Increasing the number of photonic rings (value of k) proportionally improves both the power and energy-delay characteristics of the on-chip communication infrastructure (at the cost of photonic layer complexity) due to more easily accessible photonic paths across the chip, and more efficient photonic path utilization. Thus hybrid photonic NoC fabrics like *UC-PHOTON* have a clear advantage over traditional 2D electrical NoC fabrics when it comes to reducing power and energy-delay product in future CMP designs.

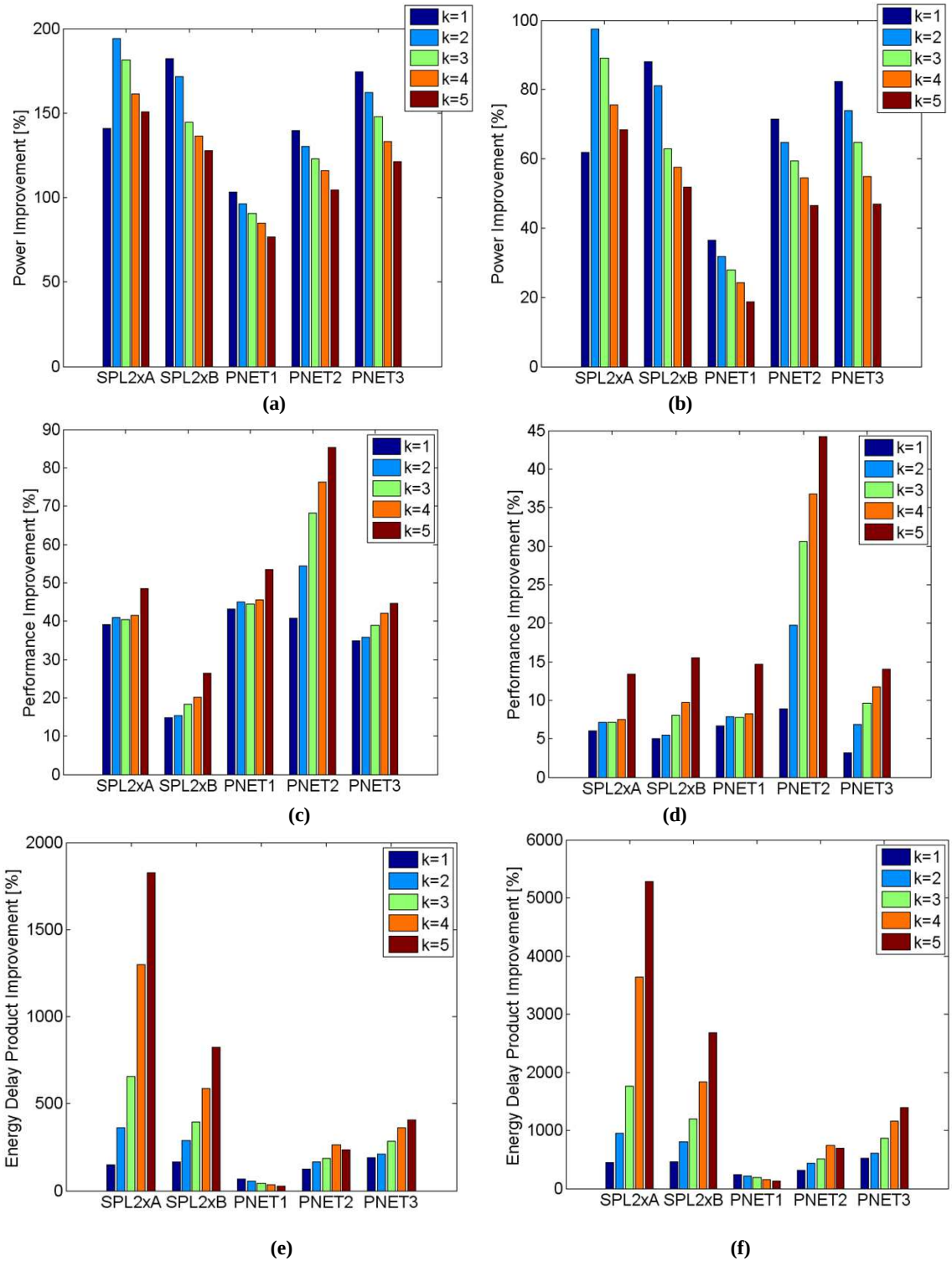


Figure 39 % Improvement for runtime reconfigurable *UC-PHOTON* vs. approaches from [120] and [119], (a) avg. power w.r.t. [120] , (b) avg. power w.r.t. [119], (c) performance w.r.t. [120], (d) performance w.r.t. [119], (e) energy-delay product w.r.t. [120], (f) energy-delay product w.r.t. [119]

In the next set of experiments, we were interested in exploring the impact of enabling runtime optimizations. Figure 38 (a)-(b) show the improvements in average power and energy-delay product for the case when runtime reconfiguration optimizations are enabled compared to the case when runtime reconfiguration options are not utilized in *UC-PHOTON*. Results are shown for the same application implementation scenarios as in the previous set of experiments, with implementations represented by bars in Figure 38 satisfying the bandwidth and latency constraints of all use-cases in the respective applications. It can be seen that enabling runtime reconfiguration has a notable impact on reducing power dissipation (*from 2.7× to 3.6×*) as well as energy-delay product (*from 1.2× to 3.1×).* For smaller benchmarks like *SPL2xA* and *SPL2xB*, an interesting phenomenon can be noticed – increasing the number of photonic rings can significantly reduce the improvements in energy-delay product. Even for larger and more complex benchmarks such as *PNET1*, *PNET2*, and *PNET3*, the improvements in energy-delay product saturate after a point and subsequently increasing the number photonic rings reduces the energy-product improvements. This happens because all runtime reconfiguration optimizations have a latency overhead (and also a small energy overhead) associated with their implementations. For more complex configurations (e.g., k = 4, 5) the latency overhead of runtime optimizations such as dynamic WDM and PRI reconfiguration can easily increase average packet latency, and consequently the energy-delay product. Overall however, it can be seen from the results that there is always a benefit in performing runtime reconfiguration, especially for power dissipation and energy-delay product.

Our final set of experiments compares the runtime reconfiguration strategies enabled in the *UC-PHOTON* architecture with the approaches proposed in the only two NoC-based works for coping with multiple use-case applications that have been previously proposed in literature. The

first approach [120] proposes creating a synthetic worst case use-case and performing runtime reconfigurations that include DVS/DFS, adaptive TDMA slot allocation, and adaptive routing during use-cases transitions.

The second approach [119], also proposes the same runtime reconfigurations, but does not create a worst-case use-case; instead optimizing on a per-use-case basis. In the interest of fairness for the comparison study, we adapt these runtime reconfiguration approaches and implement them on hybrid photonic NoC topologies similar to *UC-PHOTON*. This ensures that we are comparing the effectiveness of the runtime reconfiguration strategies.

Figure 39 (a)-(f) compares our runtime reconfiguration enabled *UC-PHOTON* architecture with the runtime reconfiguration approaches from [119] and [120] implemented on hybrid photonic NoCs. The figures show % improvement for the reconfigurable *UC-PHOTON* architecture in average power dissipation (compared to (a) approach from [120], (b) approach from m [119]), average throughput performance (compared to (c) approach from [120], (d) approach from m [119]), and energy-delay product (compared to (e) approach from [120], (f) approach from m [119]). Note that as in previous experiments, the implementations represented as bars in the figures satisfy bandwidth and latency constraints for all application use-cases. It can be seen that the reconfiguration strategies proposed with *UC-PHOTON* result in a reduction in power dissipation and energy-delay product, while also providing higher throughput performance, compared to approaches from both [120] and m [119]. These results highlight the effectiveness of the runtime PRI reconfiguration, dynamic WDM, and clock gating optimizations that are enabled only in our proposed *UC-PHOTON* communication architecture, and not proposed in the approaches from [120] and m [119].

Ultimately, the multi-ring *UC-PHOTON* configurations provide much lower average power and energy-delay products compared to single-ring configurations, but at the cost of increased complexity in the photonic layer and more overhead in the electrical layer due to the additional gateway interfaces. Our analysis of the area overhead of the different *UC-PHOTON* configurations in the electrical layer indicate that the absolute area overhead due to router enhancements in the electrical layer to implement dynamic reconfiguration schemes, deadlock recovery, and photonic interfaces increases with core count and as the number of photonic rings is increased, but is still minimal, at less than 2% chip area. In the future, as electrical NoC power dissipation is expected to be higher by a factor of around 10× than what is needed to enable tera- and petaflop performance levels of future CMPs [7] [153], innovative on-chip communication paradigms are sorely needed. Reconfigurable hybrid photonic NoC architectures like *UC-PHOTON* proposed in this work can enable up to 46× reduction in power dissipation and up to 170× reduction in energy-delay product compared to traditional electrical NoC fabrics, while still satisfying bandwidth and latency constraints for all application use-cases. This is a very promising result that motivates the need for hybrid photonic NoCs to enable high performance-per-watt communication infrastructures in future CMP architectures that implement multiple use-case applications.

4.5 RESULT SUMMARY

Emerging CMP applications today have tens to hundreds of cores and numerous use-cases with different communication bandwidth and latency requirements. Designing an on-chip communication fabric for these large systems that can satisfy the requirements of multiple use-cases at runtime is a challenging problem facing system designers today. While networks-on-chip

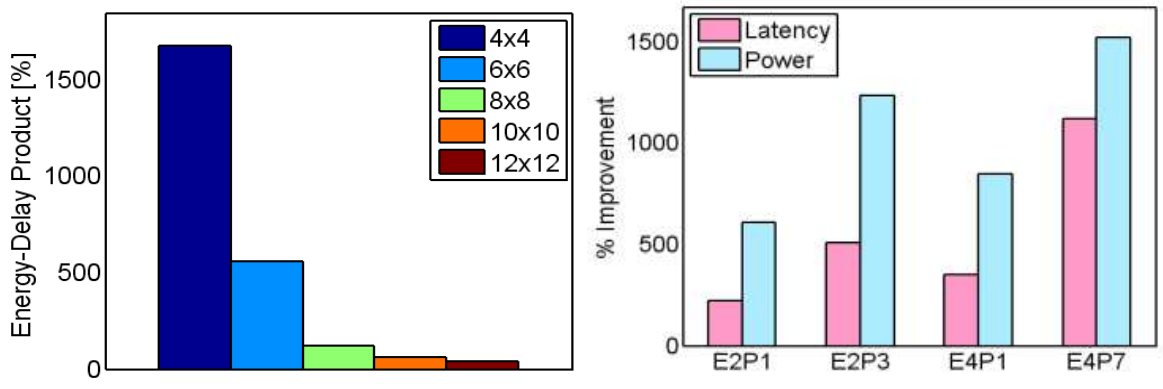
(NoCs) are a promising solution that can provide scalable performance for large CMP designs, their performance-per-watt characteristics are not satisfactory for future CMP designs. In this chapter we proposed a novel hybrid photonic NoC communication architecture called *UC-PHOTON* that can achieve scalable performance and performance-per-watt characteristics for small as well as large CMP designs. Our proposed hybrid photonic NoC architecture utilizes photonic ring paths interfaced with an electrical mesh NoC, and provides low-latency, high bandwidth, and power efficient data transfers. The novel hybrid architecture also supports various runtime reconfiguration optimizations to adapt to changing use-case performance requirements. Results of our experiments indicate that *UC-PHOTON* provides significantly better power, performance, and energy-delay characteristics compared to traditional 2D electrical NoCs designed to cope with multiple use-case applications.

5 *OPAL:* A MULTI-LAYER HYBRID PHOTONIC NOC FOR 3D ICS

Three-dimensional integrated circuits (3D ICs) offer a significant opportunity to enhance the performance of emerging chip multiprocessors (CMPs) using high density stacked device integration and shorter through silicon via (TSV) interconnects that can alleviate some of the problems associated with interconnect scaling. In this chapter we propose and explore a novel multi-layer hybrid nanophotonic-electric NoC fabric (*OPAL*) for 3D ICs. Our proposed hybrid photonic 3D NoC combines low cost photonic rings on multiple photonic layers with a 3D mesh NoC in active layers to significantly reduce on-chip communication power dissipation and packet latency. *OPAL* also supports dynamic reconfiguration to adapt to changing runtime traffic requirements, and uncover further opportunities for reduction in power dissipation. Our experimental results and comparisons with traditional 2D NoCs, 3D NoCs, and previously proposed hybrid photonic NoCs (photonic Torus, Corona, Firefly) indicate a strong motivation for considering *OPAL* for future 3D ICs as it can provide orders of magnitude reduction in power dissipation and packet latencies.

5.1 MOTIVATION FOR MULTIPLE PHOTONIC LAYERS IN 3D ICS

In general, 2D hybrid electro-photonic NoCs have an active layer with processor and memory cores interconnected using an electrical NoC interfaced to a separate silicon photonics layer consisting of photonic waveguide-based interconnect paths. In 3D ICs, multiple active layers exist and a hybrid electro-photonic NoC for 3D ICs can utilize a single photonic layer, or multiple photonic layers.



(a) **(b)**

Figure 40 Percentage improvement in (a) energy-delay product for hybrid photonic ring NoC vs. 2D electrical mesh NoC, with scaling core count, (b) average latency and power for E2P1, E2P3, E4P1, E4P7 vs. E1P1

A *single photonic layer* has the lowest design complexity, but may lack scalability. For instance, if a hybrid electro-photonic torus topology [38] is extended to 3D ICs with many more cores, a single photonic torus layer will need to be modified by increasing number of waveguides (and thus resonators, photodetectors etc) to satisfy higher bandwidth requirements from cores in multiple active layers. Not only may this not be feasible due to waveguide spacing and layout constraints, but the ensuing wider waveguide crossing losses will be prohibitively high, leading to very high laser and photonic component power dissipation [9]. Using simpler topologies such as a photonic ring [44] can be beneficial as they do not possess any crossing losses. However, a single photonic ring does not scale well when the number of cores is increased. Figure 40 (a) shows the percentage improvement in energy-delay product for a hybrid ring-mesh NoC (with a photonic ring interfaced to an electrical mesh NoC) compared to a conventional 2D electrical mesh NoC, with increasing CMP core counts. For each CMP configuration, results were averaged for various *SPLASH-2* benchmark [135] implementations. It can be clearly seen that with rising core counts, the benefits of using a single photonic ring become insignificant. This is

primarily because the photonic ring is under-utilized due to a limited number of uplinks/downlinks and coverage, even though the global communication requirements are higher.

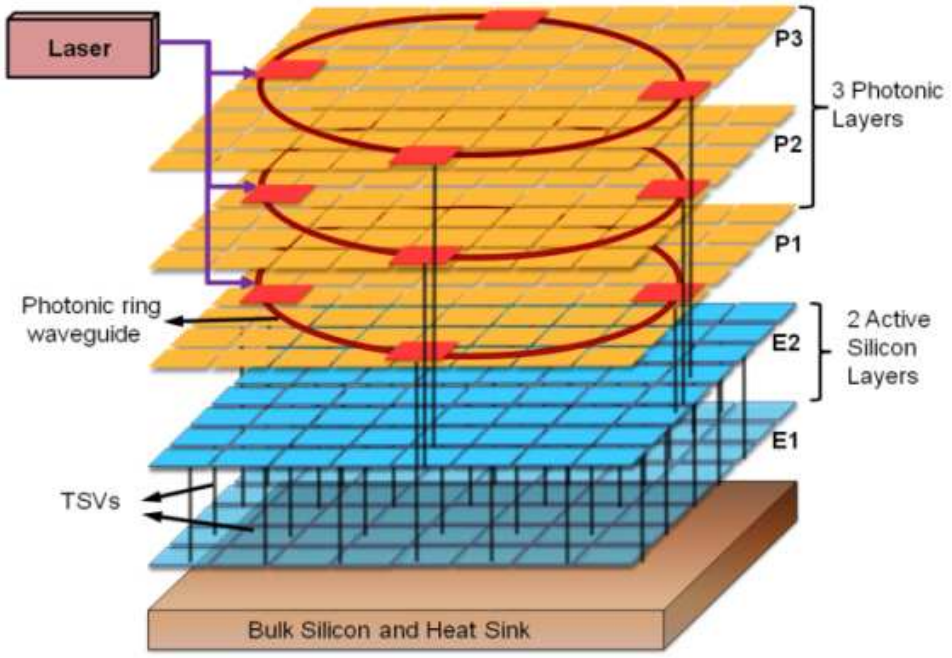


Figure 41 E2P3 *OPAL* configuration

One way to improve scalability for the hybrid ring-mesh NoC is to utilize 3D ICs with *multiple photonic layers*. For the same number of cores, a 3D IC has a smaller die area, which can enable improved coverage for the photonic ring for intra-layer transfers. In addition, dedicated photonic rings can be used to also enable inter-layer global transfers. To validate our conjecture, we performed a feasibility study to determine whether having multiple photonic layers is beneficial in 3D ICs. Figure 40 (b) shows the results of a comparison study for a hybrid ring-mesh NoC for a 100 core CMP with the following configurations: *(i)* two active layers, with 50 cores/layer and one photonic ring layer (E2P1), *(ii)* two active layers, with 50 cores/layer and three photonic ring layers (E2P3; Figure 41), *(iii)* four active layers, with 25 cores/layer and one

photonic ring layer (E4P1), and *(iv)* four active layers, with 25 cores/layer and seven photonic ring layer (E4P7). For configurations with multiple photonic layers, each layer has a dedicated photonic ring layer for intra-layer transfers, and another photonic ring layer for inter-layer transfers. Figure 42 (b) shows the percentage improvement in average power and average packet latency compared to a 100 core CMP with a single photonic ring layer and a single active layer with a mesh NoC (E1P1). A WDM degree of 32 is assumed for all configurations. It can be seen that 3D IC configurations with single photonic layers (E2P1, E4P1) provide some improvements over the E1P1 configuration, primarily due to smaller inter-layer links between cores in separate layers that replace longer global links in E1P1. However, the photonic ring was found to be the bottleneck due to high levels of traffic that caused inter-core data flows to stall. The multiple photonic layer configurations (E2P3, E4P7) perform significantly better due to a greater number of photonic paths. In the following sections, we describe our multi-layer hybrid photonic NoC architecture in detail.

5.2 *OPAL* SYSTEM LEVEL ARCHITECTURE

In this work, we propose *OPAL*, which is a hybrid electro-photonic 3D NoC architecture that employs multiple active layers and multiple photonic layers with photonic ring paths in a stack. The active layers consist of cores interconnected to each other using a 3D electrical mesh NoC. The photonic layers consist of ring shaped waveguides. *Gateway interface* routers provided the connectivity between the electrical layer and the modulators and photodetectors in the photonic layer. The choice of a photonic ring topology is motivated by the goal of reducing fabrication cost and photonic component area overhead, compared to other topologies such as mesh, torii, crossbars, and fat trees.

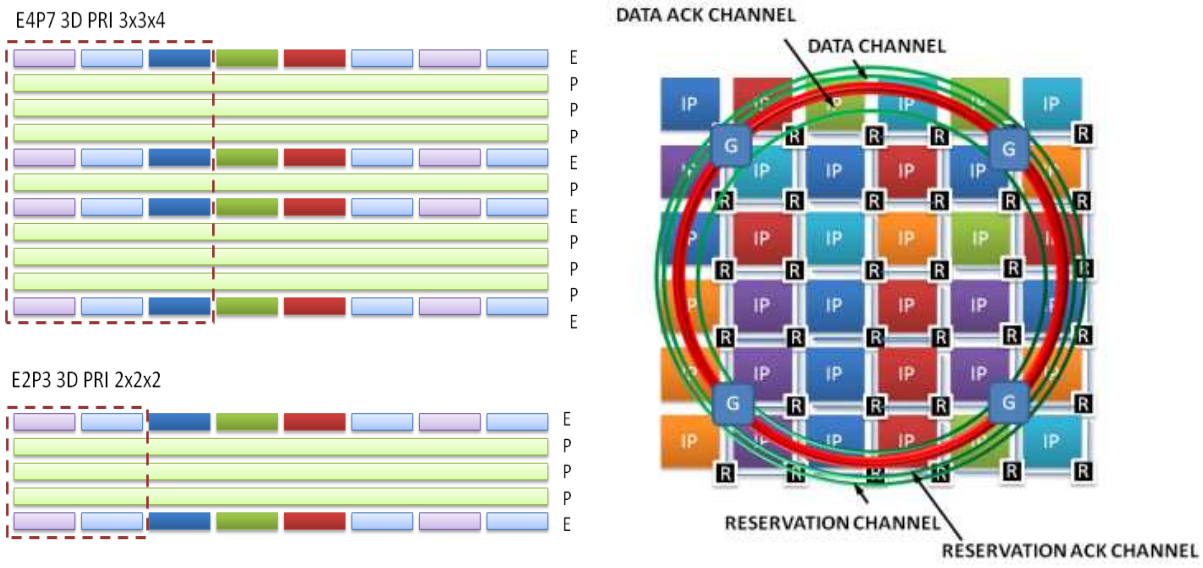


Figure 42 (a) 3D photonic region of influence (3D-PRI) (b) photonic channels

Figure 42 shows an example of a two active layer 3D IC modified to create a hybrid 3D photonic-electric network. This E2P3 *OPAL* configuration has two active electrical (E) layers and three photonic (P) layers. Each E layer has a dedicated P layer with photonic rings for intra layer global transfers between cores in the same layer. For every two E layers, a dedicated P layer exists that facilitates inter-layer (e.g. E1 to E2) global transfers. Vertical TSVs are used for transfers between E1 and E2 in the electrical 3D mesh NoC, as well as to transfer data between photonic layers and active layers. Higher complexity *OPAL* configurations can be created by reusing this basic E2P3 configuration. For instance, for a four active layer 3D IC, an E4P7 *OPAL* configuration is created by stacking two E2P3 stacks and adding a dedicated P layer for inter E2P3 photonic communication. Throughout this chapter we focus on two and four active layer 3D ICs when exploring *OPAL*, although the architecture is applicable to 3D ICs with a greater number of layers as well.

5.3 3D PHOTONIC REGION OF INFLUENCE (3D-PRI)

To balance traffic between the photonic rings and the electrical NoC, *OPAL* has a 3D parameterizable photonic region of influence (3D-PRI) which refers to the number of cores around the gateway interface that can utilize the photonic path for communication. Changing the PRI sizes can have a notable impact on communication power, latency, and bandwidth. For smaller sized systems (e.g., 2 layer, 3×3 cores/layer 3D CMPs), limiting the number of cores interfacing with each gateway interface to one may be sufficient to offload a majority of the global communication from the electrical network. However for more complex systems (e.g., 4 layer, 10×10 cores/layer 3D CMPs) a larger region size may be more appropriate. Figure 42 (a) shows examples of 3D PRIs for two *OPAL* configurations (E2P3 and E4P7). The 3D PRI for the E2P3 configuration has a size 4, which specifies 3D blocks containing 8 cores (2×2×2 – i.e., 4 cores/layer in 2 layers) around gateway interfaces that are allowed to use the photonic waveguide for transfers. For the E4P7 configuration, the 3D PRI shown has a size 9 and consists of 36 cores (3×3×4).

5.4 ROUTER ARCHITECTURE

Data flits in the *OPAL* network are transferred using wormhole switching, with flit width = 256 bits. There are broadly two types of electrical layer routers used in our proposed architecture: *(i)* electrical mesh routers that can have up to seven I/O ports (N, S, E, W, up, down, local core) and facilitate intra- and inter-layer transfers on the 3D electrical mesh NoC, and *(ii)* gateway interface routers that have one or more additional photonic interface ports and are responsible for sending/receiving flits to/from photonic interconnects in the photonic layers. As each photonic interface port has access to λ/n wavelengths for transmission (where λ is WDM

degree), we have λ/n buffers for sending data. Although it is theoretically possible to have $(n-1).\lambda/n$ data flows received at a gateway interface, we restrict the number of received flows (and hence receive buffers) to λ/n to maintain symmetry and reduce cost. All of the photonic ports are connected to a 'WDM control' module that controls wavelength assignment to different traffic flows, to enable WDM for high bandwidth photonic communication. To reduce the overhead on router complexity, only a few routers (four in our initial baseline configuration) are chosen as gateway interface routers in each active layer. To support flexible 3D-PRI sizes at runtime, each router has a region validation unit with tables that contain region boundary coordinates. Details of this, along with an overview of routing and flow control mechanisms in *OPAL* are presented next.

5.5 ROUTING AND FLOW CONTROL

To route flits in *OPAL*, a deadlock-free XYZ dimension order routing scheme is used in the electrical 3D NoC, and a modified PRI-aware XYZ routing scheme is employed for selective data transmission through the photonic links. Communicating cores lying within the same 3D-PRI communicate using the electrical NoC (i.e., intra-PRI transfers using TSVs and horizontal links). Cores that need to communicate and reside in separate PRIs communicate using the photonic paths (inter-PRI transfers), provided they satisfy two criteria: *(i)* the size of the data to be transferred is above a user-defined threshold M_{th}, and *(ii)* the number of hops from the source core to its closest PRI gateway interface is less than the number of hops to its destination core. In this way, large data messages can be offloaded from the electrical NoC and sent over a faster, more energy efficient photonic path. In addition, local communication can be done quickly via

the electrical NoC without going through expensive electrical-to-photonic and photonic-to-electrical conversions.

Transfers between cores lying outside photonic regions of influence occur normally via the electrical network using XYZ routing. Network interfaces (NIs) ensure that header flits contain coordinates of the source and destination of the packet being injected into the NoC, as well as a flag indicating that the message size is large enough to traverse a photonic path (for inter-PRI transfers). All routers in the *OPAL* architecture have region validation units that select XYZ routing for intra-PRI transfers, for transfers to cores not residing in any PRIs, or of the two photonic path criteria listed above are not satisfied. Otherwise if an inter-PRI transfer is detected by the region validation unit at the router connected to the source NI, the flits are re-routed to the gateway interface of the closest PRI using XYZ routing, traverse the photonic ring to the destination gateway interface, and then are routed to the destination core, again using XYZ routing. If multiple requests contend for access to the photonic waveguide at a gateway interface, then the request with the farthest distance to the destination is given priority.

The photonic waveguides in *OPAL* are logically partitioned into four channels: reservation, reservation acknowledge, data, and data acknowledge. In order to reserve a photonic path for a data transfer, *OPAL* utilizes a Single Writer Multiple Reader (SWMR) configuration on dedicated reservation channel waveguides. Each gateway interface has a subset of λ/n wavelengths available for transmission, where λ is the total number of wavelengths available from the multi-wavelength laser and n is the number of gateway interfaces. Every gateway interface must be able to receive $(n-1).\lambda/n$ wavelengths (from the rest of the gateway interfaces), each with a separate microring resonator receiver. A source gateway interface uses one of its available wavelengths (λ_t) to multicast the destination ID via the reservation channel to other gateway

interfaces. Each gateway interface has $\lceil log(n-1) \rceil$ dedicated SWMR reservation photonic waveguides that it writes the destination ID to, after which the other gateway interfaces read the request. Only the intended destination gateway interface accepts the request, while others ignore it. As each gateway interface has a dedicated set of λ/n wavelengths allocated to it, the destination can determine the source of the request, without the sender needing to send its ID with the multicast.

If the request can be serviced by the available wavelength and buffer resources at the destination, a reservation acknowledgement is sent back via the reservation ACK channel on an available wavelength. The reservation ACK channel also has a SWMR configuration, but a single waveguide per gateway interface is sufficient to indicate the success or failure of the request. Once the photonic path has been reserved in this manner, data transfer proceeds on the data channel, which has a low cost Multiple Writer Multiple Reader (MWMR) configuration, unlike the high overhead of several Multiple Writer Single Reader (MWSR) data channels used in Corona [40] and Firefly [37]. In *OPAL*, the number of data channel waveguides is equal to the chosen flit width (i.e., 256). The same wavelength (λ_t) used for the reservation phase is used by the source to send data on. The destination gateway interface tunes one of its available microring resonators to receive data from the sender on that wavelength after the reservation phase. Once data transmission has completed, an acknowledgement is sent back from the destination to the source gateway interface via a data ACK channel that also has a SWMR configuration with a single waveguide per gateway interface to indicate if the data transfer completed with success. Thus the overall reservation process takes a single cycle each for the path request and ACK phases at the beginning of the transfer, and one cycle for the data ACK at the end of transmission.

The advantage of having a fully photonic path setup and ACK/NACK flow control in *OPAL* is that it avoids using the electrical network for path setup, as is proposed with some other approaches [38] [44] [160], which our analysis shows can be a major latency and power bottleneck to the point of mitigating the advantage of having fast and low power photonic paths. Allowing gateway interfaces to request for access to the photonic paths whenever data is available is also more efficient than using a token ring scheme, which can suffer from low throughput and high latencies, especially under low traffic conditions [40].

5.6 DEADLOCK AVOIDANCE

While XYZ routing has been proven to be deadlock-free for mesh-like regular 3D NoCs (as no channel dependency cycles can be formed between dimensions), the modifications made to this routing scheme to accommodate photonic transfers in *OPAL* may end up creating deadlock conditions. We extensively studied deadlocks in the proposed architecture when packets traverse the photonic ring paths. To overcome a potential deadlock, we arrived at using low overhead timeout flits sporadically interleaved with the flits for the long data messages traversing the photonic paths. This is a form of regressive deadlock recovery [125]. If a timeout flit reaches a router where flits are blocked, a ‘timeout monitor’ module in the router can detect a timeout event and recognize potential cases where flits are blocked due to deadlock, and drop the blocked flits, while sending a NACK signal in the reverse direction to indicate the flits being dropped. This allows the system to unblock and recover from potential deadlock. While the method has the overhead of the additional flits in long messages intended for photonic links and a monitoring module in the routers, this is still simpler than other potential deadlock resolution alternatives

such as keeping reserved deadlock free escape channels in every router and draining deadlocked packets through the escape channels until the deadlock condition clears.

5.7 RUNTIME OPTIMIZATIONS

OPAL supports runtime dynamic reconfiguration as a way to optimize power dissipation while meeting application throughput and latency constraints. There are three primary ways in which *OPAL* enables runtime reconfiguration:

5.7.1 DVS/DFS

Dynamic supply voltage and clock frequency scaling is used during periods when performance demand is low to scale down operating voltage for the communication network to save power. *OPAL* uses a conservative model for voltage scaling, where it is assumed that the square of the voltage scales linearly with the frequency [157].

5.7.2 DYNAMIC WDM

Wavelength division multiplexing allows several photonic signals to be transmitted simultaneously in a single photonic waveguide using different wavelengths which do not interfere with each other. *OPAL* supports varying the number of WDM channels in waveguides at runtime, by shutting off channels (modulators/receivers) when data bandwidth requirements are low to save power, and enabling the channels when bandwidth requirements become high, to maintain performance goals.

5.7.3 3D-PRI RECONFIGURATION

Small PRI region sizes promote more transfers via the electrical 3D NoC, while large region sizes increase the traffic flows eligible for transfer via the photonic rings. *OPAL* supports varying the PRI size at runtime to adapt changing application traffic requirements and achieve low power operation. The reconfiguration step involves updating region boundary coordinates in tables in the *region validation* units of the NoC routers. The update phase generally lasts a few hundred cycles, during which flit injection is not allowed to maintain consistency.

5.8 EXPERIMENTS

5.8.1 EXPERIMENTAL SETUP

Photonic waveguides enable faster signal propagation compared to electrical interconnects because they do not suffer from RLC impedances. But in order to exploit the propagation speed advantage of photonic interconnects, electrical signals must be converted into light and then back into an electrical signal. This process requires a performance and power overhead that must be taken into account for an accurate analysis. To explore the impact of using *OPAL* in CMPs, we modeled *OPAL* by extensively modifying our in-house cycle accurate SystemC-based NoC simulator. Six benchmarks from the *SPLASH-2* suite [135] were selected (*Cholesky, FFT, Fmm, Lu, Radix, Ocean*), parallelized, and implemented on multiple cores in the simulator model.

We targeted a 32 nm process technology, and assumed a fixed 400 mm^2 CMP active die area budget. Thus a single active layer 2D NoC configuration has a 400 mm^2 active E layer die area, an E2P3 configuration has a 200 mm^2 die area per active E layer, and an E4P7 configuration has a 100 mm^2 die area per active E layer. The operating frequency of the photonic rings was estimated by calculating the time needed for the light to travel from any node to the

farthest node, so that data can be transmitted to all nodes in one cycle. Through geometric calculations for the rings, and incorporating latching delays (using ITRS data [6]) we obtained a maximum operating frequency of greater than 3 GHz for the different sizes of CMPs we considered. Ultimately, the photonic rings and the communication network were clocked conservatively at 2.3 GHz. The data message threshold size for inter 3D-PRI photonic transfers was fixed at 2048 bits, and the packet size was kept at 10 flits. Delay estimates for the various photonic interconnect-centric components used in *OPAL* were obtained from [89] and from device fabrication results [140]. The delay of an optimally repeated and sized electrical (Cu) wire at 32 nm was assumed to be 42ps/mm [29].

The power dissipated in *OPAL* can be categorized into two components: electrical network power and photonic ring network power. The static and dynamic power dissipation of electrical routers and links in this work is based on results from Orion 2.0 [141] incorporated into our simulator. For calculating power dissipation of the modulator driver and TIA power we used ITRS device projections [4] and standard circuit procedures. In addition, an off-chip electrical laser power of 3.3 W *per photonic layer* (with 30% laser efficiency) is also considered in the power calculations. The laser power value accounts for per component optical losses for the coupler/splitter (1.2dB), non-linearity (1dB at 30mW), waveguide (3dB/cm), waveguide crossings (0.05dB), ring modulator (1dB), receiver filter (1.5dB) and photodetector (0.1 dB).

5.9 RESULTS

5.9.1 COMPARISONS WITH 2D AND 3D ELECTRICAL MESH NOC

In the first set of experiments, we compared the performance and power characteristics of the E2P3 and E4P7 *OPAL* configurations, but without enabling any dynamic reconfiguration,

with traditional 2D and 3D electrical mesh NoCs. The 2D configurations considered included a 64 core (8×8) and 100 core (10×10) NoC, while the 3D configurations included a 2 layer 64 core (8×4×2), a 2 layer 100 core (10×5×2), a 4 layer 64 core (4×4×4), and a 4 layer 100 core (5×5×4) NoC. The *OPAL* configurations have four uplinks between an active layer and its associated photonic layer, a PRI size of two, and WDM with 32 wavelengths in the photonic waveguides.

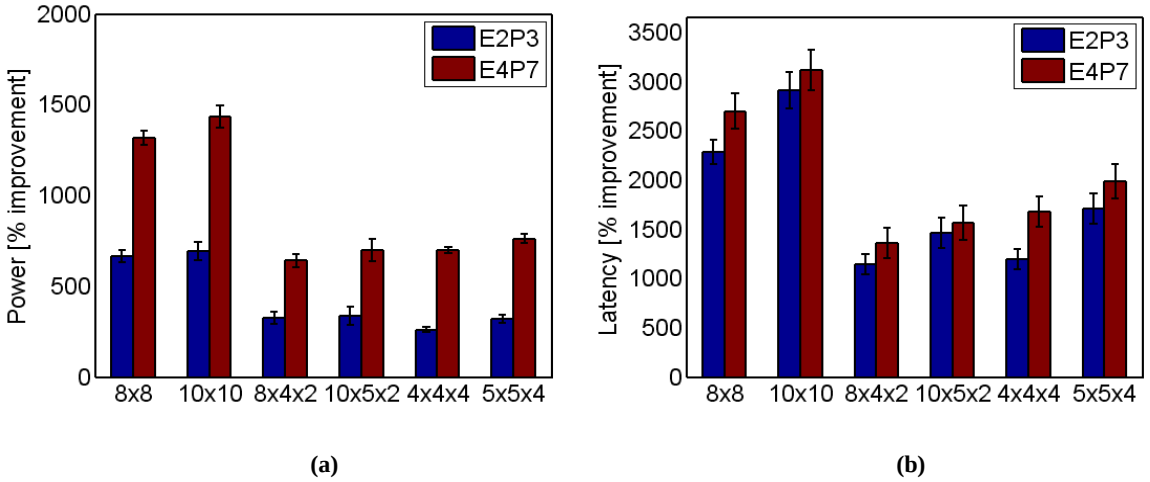


Figure 43 Percentage improvement for *OPAL* configurations compared to 2D and 3D electrical mesh NoCs (a) power, (b) average packet latency

Figure 43 (a)-(b) show the improvements in power and average packet latency for the E2P3 and E4P7 *OPAL* configurations compared to the 2D and 3D electrical mesh NoCs. It can be seen that 3D electrical mesh NoCs have a much lower power dissipation and average packet latency compared to 2D electrical mesh NoCs which explains the recent interest in 3D ICs and the potential gains that can be achieved by shifting from 2D to 3D ICs. *OPAL* goes a step farther and outperforms the all-electrical 3D ICs because of its use of low power and high speed photonic interconnects. In general, the E4P7 *OPAL* configuration outperforms the E2P3 configuration and obtains an up to a 15× power reduction and 32× average latency reduction over 2D ICs, and up

to a 7× power reduction and 19× average latency reduction over 3D ICs. These results indicate that *OPAL* has the potential to improve the benefits that can be achieved by using 3D ICs in future CMP designs.

5.9.2 IMPACT OF VARYING NUMBER OF UPLINKS

To overcome the bottleneck of a limited number of uplinks (i.e., gateway interfaces), we next explored the impact of varying the number of uplinks in the *OPAL* architecture at design time and measured the performance and power for the various configurations. As the number of gateway interface routers with photonic interfaces increases, it also results in an increase in power due to electro-photonic conversion. Increasing the number of uplinks also increases real estate usage in the silicon layer, as well as the complexity of the photonic layer. However the additional complexity of more uplinks can translate into better photonic path utilization for communication flits in some applications. In addition, increasing the number of uplinks can also provide fault tolerance in case of uplink failures. Figure 44 shows results of varying the number of uplinks for a 100 core CMP with a fixed PRI region size of four for E2P3 (2×2×2 cores/region), and E4P7 (2×2×4 cores/region), and WDM with 32 wavelengths in the photonic waveguides.

For a configuration with η uplinks, there are 2η gateway interfaces per active (E) layer for E2P3 (η interfaces to the private P layer, and η interfaces to the shared P layer), and 3η gateway interfaces per active (E) layer for E4P7 (η interfaces to the private P layer, and 2η interfaces to the two shared P layers). Improvements in power dissipation and latency were significant when the number of uplinks were increased from 4 to 8. The improvements drop when uplinks are increased from 8 to 16 due to overlapped PRI regions leading to less opportunity for global communication. This trend continues with further increase in the number of uplinks, with the 32

uplink case providing negligible improvements over the 16 uplink case, while significantly increasing complexity in the photonic and electrical NoC layers.

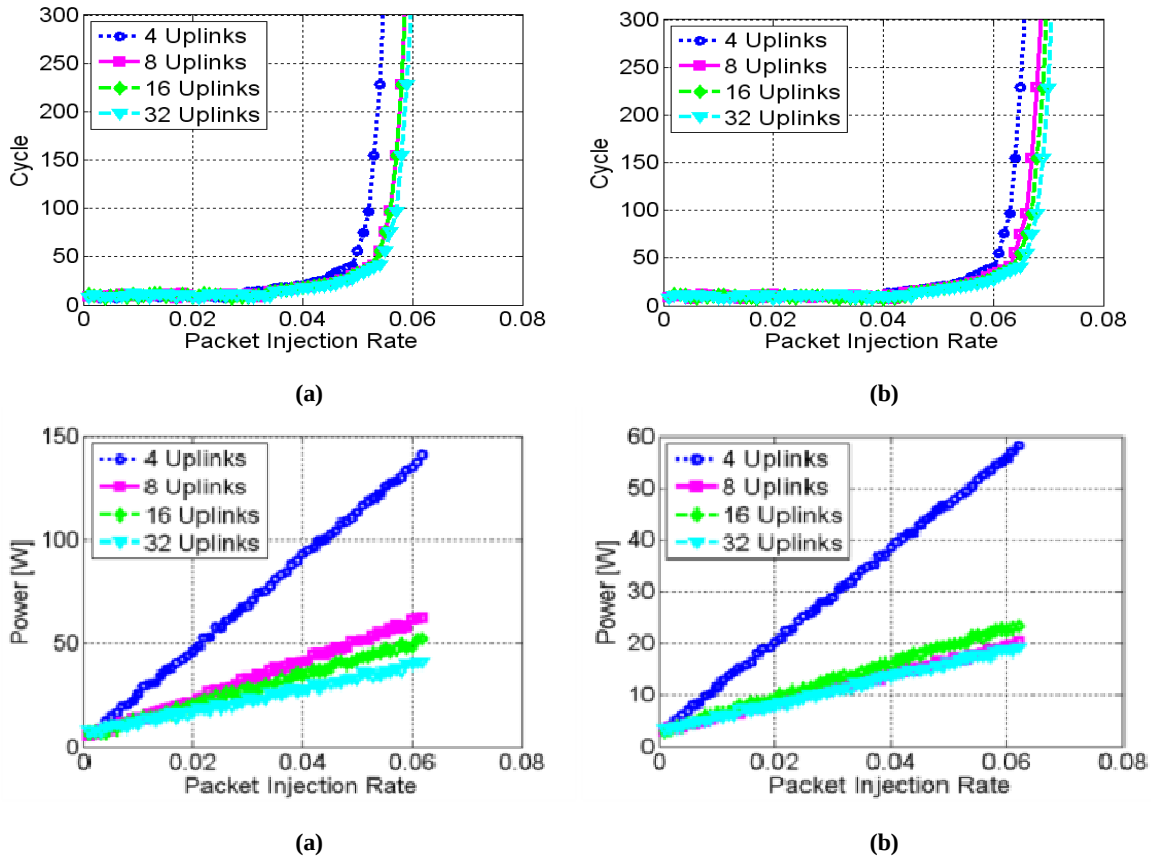


Figure 44 Impact of changing number of uplinks on (a) latency of E2P3, (b) latency of E4P7, (c) average power of E2P3, (d) average power of E4P7

5.9.3 IMPACT OF ENABLING RUNTIME ADAPTATIONS

In the next set of experiments, we explored the impact of enabling runtime adaptation in *OPAL* on the overall power dissipation. Dynamically adapting resources based on runtime traffic requirements can expose opportunities for power savings. Figure 46 presents results of power savings for the six *SPLASH-2* benchmark implementations when the dynamic reconfiguration schemes (PRI resizing, WDM scaling, DVS/DFS) are applied simultaneously, compared to the

baseline case without any dynamic reconfiguration enabled. The implementation of these schemes was guided by offline profiling of the selected benchmark implementations. Results are shown for the E2P3 and E4P7 *OPAL* configurations, for 64 and 100 core CMPs with a WDM degree of 32 and four uplinks.

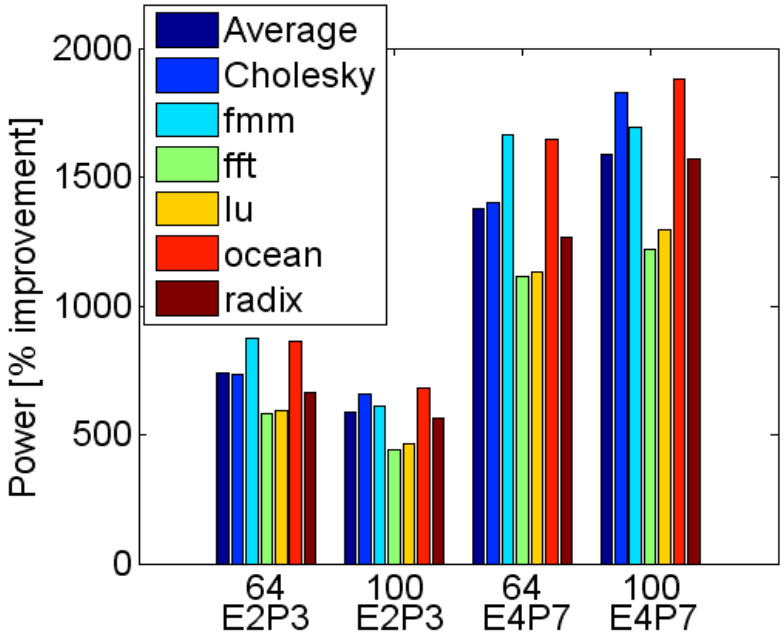


Figure 45 Percentage improvement in average power dissipation for E2P3 and E4P7 *OPAL* configurations, with all runtime adaptations enabled (DVS/DFS, WDM, PRI) relative to baseline case with no runtime adaptation enabled

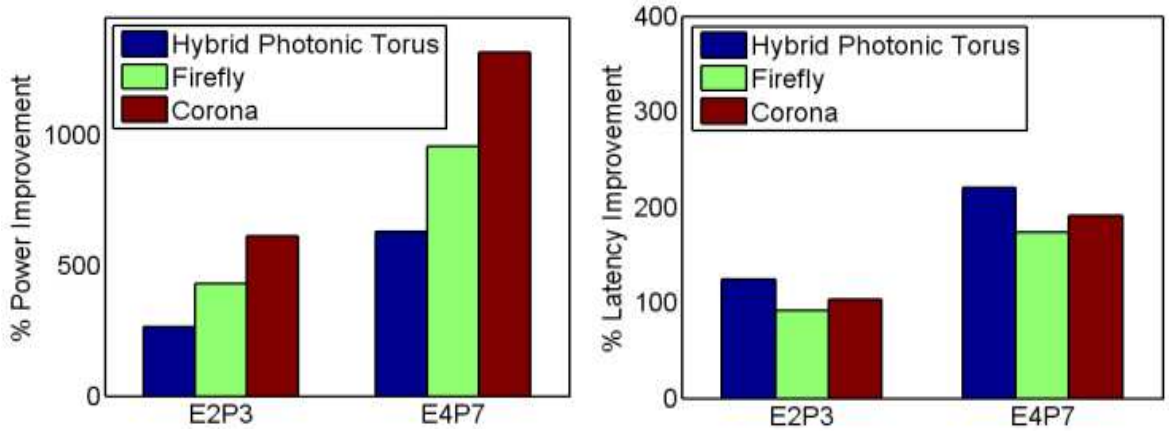
It can be seen from Figure 46 that the cumulative improvement in power savings for the optimizations is significant. It was found that the improvements due to DVS/DFS diminish with increasing core count due to the increased overhead of the DVS/DFS circuitry, and smaller sized links which provide lower power savings. A similar trend is noticed with WDM scaling, with diminishing improvements as core count increases. This is due to greater demand for photonic communication by the increased number of traffic flows which limits the opportunities for

reducing wavelength channels for WDM. As the number of active (E) and photonic (P) layers increase, the number of gateway interfaces and consequently area covered by 3D-PRI regions also increase. The E4P7 configuration therefore has more opportunities for fine tuning traffic distribution among electrical and photonic paths by utilizing PRI reconfiguration compared to the E2P3 configuration, leading to an increase in power savings. The improvements due to PRI resizing overshadow the diminishing returns from DVS/DFS and WDM scaling for the E4P7 configuration as core counts increase, which is why its power dissipation improves (reduces) with increasing core counts. The E2P3 configuration does not benefit as much by utilizing PRI resizing with increasing numbers of cores, and consequently has lower power savings for higher core counts.

5.9.4 COMPARISON WITH EXISTING HYBRID PHOTONIC NOCS

Our final set of experiments compares the E2P3 and E4P7 *OPAL* 3D hybrid photonic NoC configurations with three previously proposed 2D hybrid photonic communication architectures: (i) a hybrid photonic torus interfaced with an electrical 2D torus NoC [38], (ii) the hybrid Corona architecture [40], and (iii) the hybrid Firefly architecture [37]. Both *OPAL* configurations utilized dynamic reconfiguration and 8 uplinks. For fairness of comparison, all the compared architectures were modeled with a WDM degree of 128, and were simulated using the same set of technology parameters, component delay and power models, and traffic. Results were obtained for a 100 core CMP. Figure 46 (a)-(b) shows the percentage improvement for the E2P3 and E4P7 *OPAL* configurations in terms of power dissipation and average packet latency over the hybrid photonic torus, Corona, and Firefly architectures. From the results it can be seen that the *OPAL* configurations improve upon existing 2D hybrid photonic NoC architectures, with the

E4P7 configuration showing somewhat higher improvements than the E2P3 configuration. For instance, compared to the Firefly hybrid NoC, the E4P7 *OPAL* configuration shows up to approx. 8× reduction in power dissipation and a 3× reduction in average packet latency. The ability to better balance traffic between the electrical and photonic paths, a more effective photonic path setup, support for runtime adaptations of the electrical and photonic networks, and the use of shorter TSVs to replace longer global wires are the primary reasons for *OPAL's* superior performance.



(a) **(b)**

Figure 46 Percentage improvement for E2P3 and E4P7 *OPAL* configurations compared with hybrid photonic torus [32], Corona [28] and Firefly [29] NoCs: (a) power dissipation (b) average packet latency

In terms of photonic component area overhead, our calculations indicate that the E4P7 *OPAL* configuration has lower photonic component area by a factors of 1.5×, 1.6×, and 2.1× compared to Firefly, photonic torus, and Corona architectures respectively. we conjecture that compared to having a single complex photonic layer, having multiple simpler photonic layers as in *OPAL* can not only ease fabrication challenges, but also provide lower average power and

latency as the experimental results indicate. These results also make a strong case for considering the use of photonic interconnects in emerging 3D ICs.

5.10 RESULT SUMMARY

In this chapter, we proposed and explored a multi-layer hybrid electro-photonic NoC fabric (*OPAL*) for 3D ICs. Our proposed 3D hybrid ring-mesh NoC combines low cost photonic rings on multiple photonic layers with 3D mesh NoCs in active layers to reduce on-chip communication power dissipation and latency. *OPAL* also supports mechanisms for adaptation to changing traffic at runtime to optimize power dissipation. Experimental comparisons with traditional 2D NoCs, 3D NoCs, and previously proposed hybrid photonic NoCs indicate a strong motivation for considering *OPAL* for future 3D ICs as it can provide several orders of magnitude reduction in power dissipation and average latency.

6 SYNTHESIS FRAMEWORK FOR APPLICATION-SPECIFIC HYBRID NANOPHOTONIC-ELECTRIC NOCS WITH WAVEGUIDES

To date, prior work on automated NoC synthesis has mainly focused on electrical NoCs. In this chapter, for the first time we propose a suite of techniques for effectively synthesizing hybrid nanophotonic-electric on-chip interconnects. We formulate and solve the synthesis problem using four search heuristics: *(i)* Ant Colony Optimization (ACO), *(ii)* Particle Swarm Optimization (PSO), *(iii)* Genetic Algorithm (GA), and *(iv)* Simulated Annealing (SA).

6.1 MOTIVATION FOR HYBRID NOC SYNTHESIS

CMOS compatible on-chip photonic interconnects with silicon-on-insulator waveguides provide a potential substitute for electrical interconnects, particularly for global on-chip communication, allowing data to be transferred across a chip with much faster light signals. Based on recent technological advancements, the critical length at which photonic interconnects are advantageous over electrical interconnects has fallen to well below chip die dimensions. To minimize power, recent research [45] [46] [48] [49] [50] [51] [52] [161] [162] has focused on novel hybrid photonic NoC architectures that optimize the distribution of local and global communication between electrical and photonic links. The optimization of these hybrid photonic NoCs for parallel embedded applications requires traversing a massive design space to determine suitable application-specific values for parameters such as wavelength division multiplexing (WDM) density, number of photonic uplinks, serialization degree, etc. in order to maximize communication performance-per-watt. For example, a photonic interconnect with n=256 waveguides and m=256 wavelengths will require exploring $n+(n)^2+(n)^3+\ldots+(n)^m$ configurations to find the most power efficient solution that also meets performance goals. This is most

certainly practically exorbitant. Moreover, these two parameters are just a small subset of the much larger set of parameters that must be explored during hybrid photonic NoC optimization. Finding the best solution for such a combinatorial optimization problem that is known to be NP-hard could take years if we search through the entire solution space, even with leading-edge supercomputing technology today. Indeed, application-driven optimization of hybrid photonic NoCs will become increasingly important as on-chip core counts increase, but this problem has not yet been addressed in prior work by researchers. One viable way to solve this optimization problem for hybrid photonic NoCs is by developing polynomial-time heuristics that permit us to identify and search through a relevant portion of the solution space in a tractable amount of time to find a near optimal solution. Greedy heuristics are unlikely to find good quality solutions due to their inclination for getting stuck in local optima. In contrast, non-greedy search heuristics such as Simulated Annealing (SA) [56] operate through repeated transformations and have the hill climbing ability to escape local optima by allowing acceptance of worse solutions within the evaluation process. A population of solutions being simultaneously manipulated is one of the major differences between the SA and traditional greedy search algorithms. Approaches based on SA and other non-greedy iterative algorithms have proven highly effective in recent years for several hard problems in the realm of VLSI physical design, such as partitioning and placement [163]. In this chapter, we address the problem of synthesizing (i.e., optimizing) application-specific hybrid photonic NoCs.

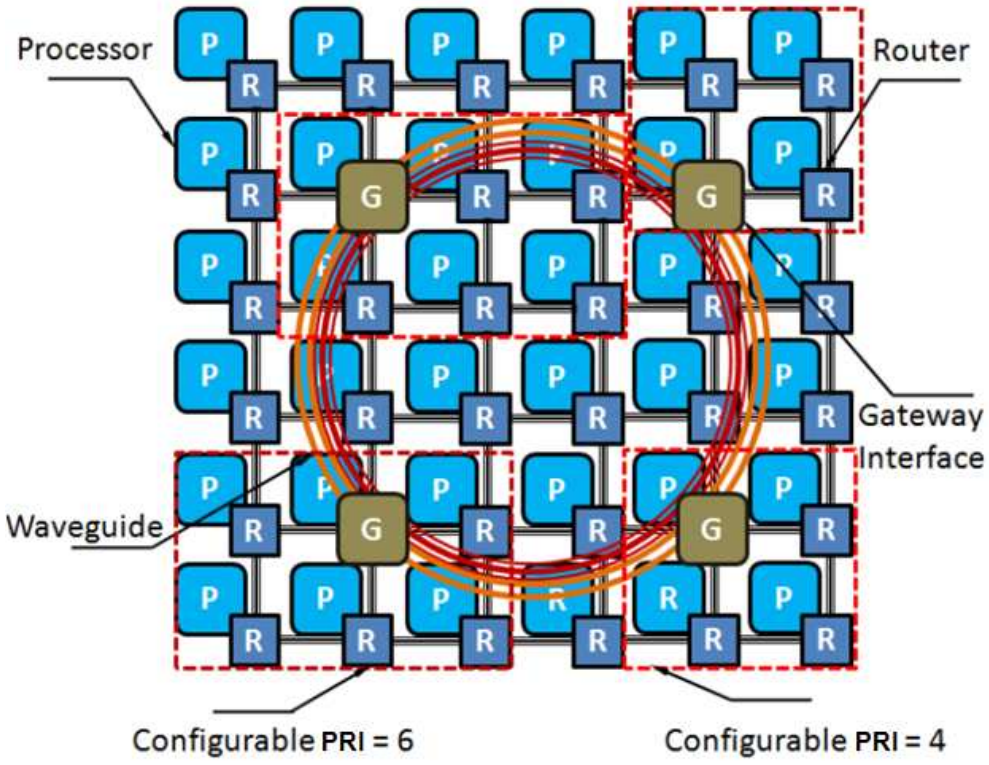


Figure 47 Hybrid ring-mesh photonic architecture [44]

6.2 HYBRID PHOTONIC NOC ARCHITECTURE OPTIMIZATION PARAMETERS

We consider the ring-mesh hybrid photonic topology presented in [44] as our baseline architecture. Here we summarize some of the key features of this hybrid photonic NoC architecture. Figure 47 shows an overview of the architecture, which consists of concentric ring photonic waveguides on a dedicated photonic layer, interfaced to a 2D electrical mesh NoC. The key motivation of this hybrid architecture is to use photonic links opportunistically to reduce latency and power dissipation for global communication while utilizing the electrical NoC for local and semi-global communication. The electrical mesh is composed of two types of routers: *(i)* conventional four stage pipelined electrical mesh routers that have 5 I/O ports (N, S, E, W, local core) with the exception of the boundary routers that have fewer I/O ports, and *(ii)* gateway

interface routers that are also four-stage pipelined but have six I/O ports (N, S, E, W, local core, photonic link interface) and are responsible for sending/receiving flits to/from photonic interconnects in the photonic layer.

6.2.1 PRI-AWARE ROUTING

A unique feature of this hybrid photonic NoC architecture is the reconfigurable traffic partitioning between the electrical and photonic links. To minimize implementation cost, the number of gateway interfaces are kept low (e.g., 4 or less). However, with increasing CMP core counts, fewer gateway interfaces reduce photonic path utilization. To ensure appropriate scaling and utilization, a parameterizable photonic region of influence (PRI) is used, which refers to the number of cores around the gateway interface that can utilize the photonic path for communication. For larger CMPs, having a larger PRI size can ensure appropriate photonic path utilization. Figure 47 shows an 8×8 (64 core) CMP with varying PRI sizes at four gateway interfaces. A modified PRI-aware XY routing scheme routes packets in this architecture as follows. Communicating cores lying within the same PRI region communicate using the electrical NoC (intra-PRI transfers). Cores that need to communicate and reside in different PRIs communicate using the photonic paths (inter-PRI transfers), provided they satisfy two criteria: *(i)* the size of data to be transferred is above a user-defined size threshold M_{th}, and *(ii)* the number of hops from the source core to its local PRI gateway interface is less than the number of hops to its destination core.

6.2.2 PHOTONIC RING CONFIGURATION

The concentric ring photonic waveguides are logically partitioned into four channels: reservation, reservation acknowledge, data, and data acknowledge. A fully photonic path setup

and acknowledgement mechanism is implemented, with the reservation and acknowledge channels utilizing a Single Writer Multiple Reader (SWMR) configuration and the data channel utilizing a low cost Multiple Writer Multiple Reader (MWMR) configuration. Each gateway interface has a subset of λ/n wavelengths (microresonator modulators) available for transmission, where λ is the total number of wavelengths available from the multi-wavelength laser and n is the number of gateway interfaces. Every gateway interface must be able to receive $(n-1)\times\lambda/n$ wavelengths (from the rest of the gateway interfaces), each with a separate microring resonator receiver. A source gateway interface uses one of its available wavelengths (λ_t) to multicast the destination ID via the reservation channel to other gateway interfaces. Each gateway interface has $\lceil log(n-1) \rceil$ dedicated SWMR reservation photonic waveguides that it writes the destination ID to, after which the other gateway interfaces read the request. Only the intended destination gateway interface accepts the request, while the others ignore it. As each gateway interface has a dedicated set of λ/n wavelengths allocated to it, the destination can determine the source of the request, without the sender needing to send its ID.

6.2.3 FLOW CONTROL

If the request can be serviced by the available wavelength and buffer resources at the destination, a reservation acknowledgement is sent back via the reservation ACK channel on an available wavelength. The reservation ACK channel also has a SWMR configuration, but a single waveguide per gateway interface is sufficient to indicate the success or failure of the request. Once the photonic path has been reserved in this manner, data transfer proceeds on the data channel, which has a low cost Multiple Writer Multiple Reader (MWMR) configuration. As flits are routed through the nearest gateway interface, global communication power consumption

is significantly lowered and the electrical network bandwidth availability is increased, enabling a win-win scenario. Once data transmission has completed, an acknowledgement is sent back from the destination to the source gateway interface via a SWMR channel, with a single waveguide per gateway interface to indicate if the data transfer successfully completed or failed. The advantage of a fully photonic path setup and ACK/NACK flow control is that it avoids using the high latency electrical network, as done in prior work (e.g., [102]). Our architecture thus allows gateway interfaces to request for access to the photonic paths whenever data is available. This scheme is more efficient than using a token ring, which can suffer from low throughput and high latencies, especially under low traffic conditions. High throughput is achieved by using dense wavelength division multiplexing (DWDM), with multiple wavelengths per waveguide available to transfer multiple streams of concurrent data.

6.2.4 SERIALIZATION

To reduce the number of photonic components (waveguides, buffers, ring resonator based transmitters/ receivers, photodetectors), and consequently reduce area and power dissipation in the photonic layer, we also make use of serialization at the gateway interfaces. We use a shift register based serialization scheme. A single serial line is used to communicate both data and control signals between the source and destination nodes. A frame of data transmitted on the serial line using this scheme consists of n+2 bits, which includes a start bit (‘1’), n bits of data, and a stop bit (‘0’). When a word is to be transferred, the ring oscillator is enabled and it generates a local clock signal that can oscillate above 2 GHz to provide high transmission bandwidth. At the first positive edge of this clock, an n+2 bit data frame is loaded in the shift register. In the next n+1 cycles, the shift register shifts out the data frame bit by bit. The stop bit

is eventually transferred on the serial line after n+2 cycles, and r0 becomes '1'. At this time, if the transmission buffer is empty, the ring oscillator and shift registers are disabled, and the serial line goes into its idle state. Otherwise, the next data word is loaded into the shift register and data transmission continues without interruption.

Table 6 Synthesis parameters

Synthesis Parameters	Range low	Range high
Photonic Uplinks	4	32
PRI	1	(num_cores)/4
WDM Density	32	256
Serialization Degree	1	32
Clock Frequency (GHz)	1	6
PRI data size threshold (M_{th})	4	1024
Flit Width (bytes)	4	256
Waveguides	2	256

6.3 PROBLEM FORMULATION

Our synthesis problem has the following inputs:

(i) A core graph *G (V, E)*; with the set *V* of vertices $\{V_1, V_2, V_3,\ldots,V_N\}$ representing the *N* cores on which the given applications tasks have already been mapped, and the set of *M* edges $\{e_1, e_2, e_3,\ldots,e_M\}$ with weights that represent application-specific latency constraints between communicating cores,

(ii) A regular mesh-based CMP with *T* tiles such that $T = (d^2)$, where *d* is the dimension of the mesh, and each tile consists of a compute core and a NoC router,

*(iii)*The upper and lower bounds that define an acceptable value range for a set of parameters relevant to hybrid photonic NoC architectures, as defined in Table 6.

Objective: Given the above inputs, our goal is to synthesize a hybrid photonic-ring/electrical-mesh NoC architecture that will determine *(i)* number and location of photonic uplinks (i.e., gateway interfaces), *(ii)* PRI sizes, *(iii)* density of wavelength division multiplexing (WDM), *(iv)* serialization degree, *(v)* link clock frequency, *(vi)* data threshold size, *(vii)* flit widths, and *(viii)* number of photonic waveguides, while satisfying the target applications communication latency constraints and optimizing (minimizing) overall communication power dissipation. We focus our synthesis efforts on regular topologies because we believe that future chips with hundreds of cores will be much more predictable in the face of process variations, easier to design, and simpler to verify if the underlying network structure is homogeneous, even if the cores themselves are heterogeneous.

6.4 SYNTHESIS FRAMEWORK OVERVIEW

In this section, we present an overview of our hybrid photonic NoC synthesis framework. Figure 48 shows a high level flow diagram of our synthesis framework that starts with a given core graph *G(V, E)* and constraints defined in Table 6. In the first step, we perform core-to-tile mapping to optimize the aggregate communication bandwidth and power in the network. The second step focuses on parametric NoC synthesis utilizing novel implementations of the four search algorithms we consider, aimed at further reducing power dissipation while satisfying latency goals. In the final step, we verify our synthesis results using a cycle-accurate SystemC simulation to account for fine-grained traffic congestion and interference effects that can only become apparent with detailed simulation analysis. The following sections present a detailed description of these three steps.

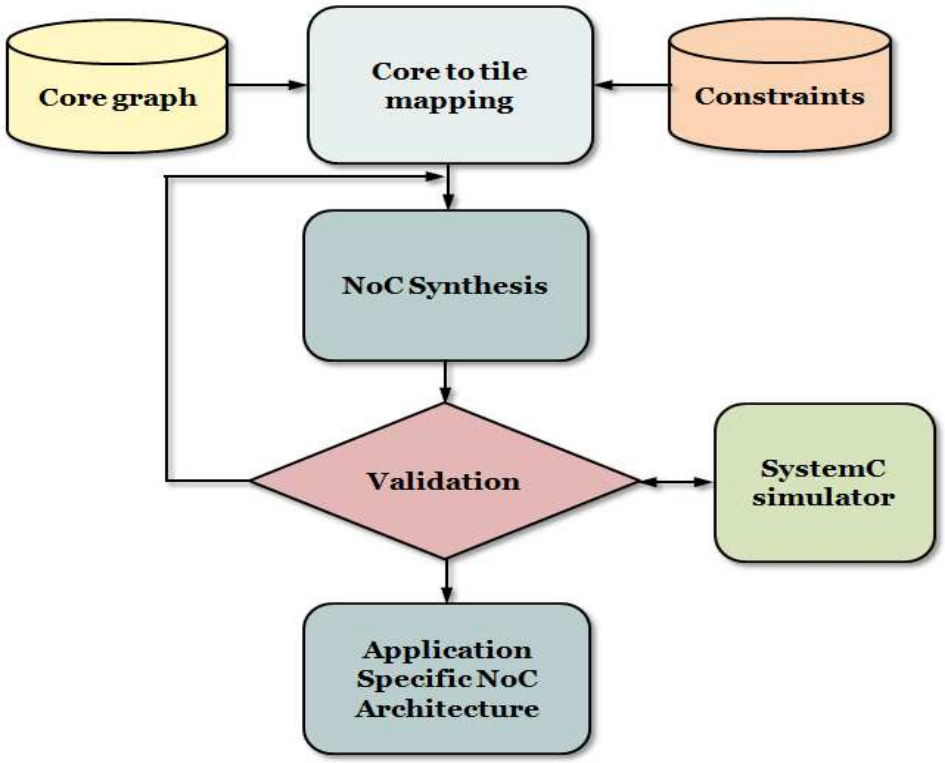


Figure 48 NoC synthesis design flow of the synthesis process

6.4.1 CORE TO TILE MAPPING

We performed one-to-one core-to-tile mapping by enhancing [45] 2D electrical NoC approach to the hybrid nanophotonic-electric architecture. To perform core-to-tile mapping we developed a greedy heuristics that minimizes communication work load $\psi_i = \sum_{v=1}^{n}[p_n \times w_n \times \emptyset_n \,]for \,\forall\, V_n$, where p_n is defined as the number packets communicated from a source node i to all its n destination nodes, w_n is defined as a weight representing power, and $\emptyset_n$ is defined as Manhattan distance. Figure 49 shows pseudo-code for our greedy heuristics, *ComputeWorkload()* computes ψ_i for all *i* cores. Then we ranked ψ_i for all *i* cores in descending order by *RankWorkLoad()* function. Core with highest communication workloads ψ_i were assigned directly or within proximity to the gateway interfaces by *AssignUplink()* achieving

communication workload reduction. This approach provided 20% average initial latency reduction compared to the random mapping.

```
done = 1
while(done)
      for (i=0; i++; i< CORES)
         ComputeWorkload();       // compute workload Ψ for each core
      end for  // end of the loop
      RankWorkLoad();             // rank the workload
      for (i=0; i++; i< Uplinks)
         AssignUplink();           // Assign Uplink to high
     end for                       // communication cost cores
    done = 1
end while
```

Figure 49 Core-to-tile mapping greedy heuristics

6.4.2 NOC SYNTHESIS

In this subsection, we present details of each of the four search heuristics based on Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Simulated Annealing (SA), and Genetic Algorithm (GA) that we utilize to perform hybrid photonic NoC synthesis.

6.4.2.1 PARTICLE SWARM OPTIMIZATION (PSO)

The Particle Swarm Optimization (PSO) metaheuristic was initially proposed by R. Eberhart and J. Kenned [164] in 1995. The fundamental idea behind PSO is inspired by the coordinated and collective social behavior of species like a flock of birds, fish, termites, or even humans. In nature, each individual bird, bee, or fish shares some information with its neighbors

and by utilizing shared information collectively, they strive to organize efforts such as developing flying patterns to minimize aerodynamic drag, etc. Although by itself, a single entity such as a bird or a bee is a simple and unsophisticated creature, collectively as part of a swarm they can perform complex and useful tasks such as building nests, and foraging. Within the PSO framework, an individual entity is called a particle and it shares information with other entities, either in the form of direct or indirect communication to coordinate their problem-solving activities. In recent years, the PSO algorithm has been applied to many combinatorial optimization problems such as optimal placement of wavelength converters in WDM networks [165] and dynamic reconfiguration of field-programmable analog circuits [166].

To implement the PSO algorithm, particles are placed in the search space of some problem, and each particle evaluates the objective function at its current location to determine its next movement by combining the best (best-fitness) locations in the vicinity. The next iteration takes place after all particles are relocated to the new position. This process is repeated for all particles and eventually for the swarm as a whole; similar to the flock of birds collectively foraging for food. A particle on its own does not have power to solve the problem; rather the solution evolves as the particles interact and work together, utilizing a social network consists of bidirectional communication. The movement of each particle is affected by its inertia or own weight and directional velocity towards local and global best solutions. As the algorithm iterates, particles move towards local as well as global solution optima forming a swarm pattern. For example, when one particle or entity finds a good solution such as a food source, other particles are more likely to be attracted by following a positional path. This social interaction feedback eventually causes all particles to move towards a globally optimal solution path. The particles search or move in the solution space by gravitating towards optimality based on the neighborhood and

global particle fitness. This transversal phenomenon is similar to the social interactions where a leader or a set of leaders emerge from the swarm and followers attempt to follow them. In summary, the idea of the PSO is to mimic the social collective behavior found in nature and utilizing it to solve complex problems.

```
done = 1
while(done)
    for (i=0; i++; i< N_ITER)
   InitializeParticles();    // generate m  particles
   UpdateParticleSystem();  // update particle system for local  and global best solution
                            //move  the particles to new position
    UpdatePositionMatrix();// position update for each particle
     end for // end of the loop
    if Termination criterion met then
       done = 0
    else
       done = 1  //Continue with next PSO iteration
    end if

end while
```

Figure 50 Particle swarm optimization formulation

Figure 50 shows the pseudo-code for our PSO formulation. The algorithm starts by initializing each particle with the function call *InitializeParticles()*. This function initializes inertia and learning weights, initial position and velocity for each parameter (from Table 6) such as WDM, PRI, Waveguide etc. For the PRI, it also generates source and destination locations of

the communication neighborhoods based on the application communication trace. The function *UpdateParticleSystem()* iterates and updates velocity and positions of the individual particles, using relations *(1)* and *(2)* that are presented later in this section. At the end of the evaluation loop, particle positions are updated and they are moved to new positions by calling the function *UpdatePositionMatrix().*This process continues for the entire application, until a dominating solution emerges.

A communication request or flit represents a particle in our PSO synthesis process. A group of random particles or random solutions are initialized during the initial phase of PSO and then the optimal or near optimal solution is constructed using an iterative synthesis process. The flits are routed from the source to the destination in a NoC. The PSO process can select among various values for the parameters in Table 6, such as number of WDM channels, or PRI size. As more core communications are considered, based on relations *(1)* and *(2)* gradually one dominant solution emerges. This best configuration satisfies application-specific latency constraints and has the lowest power dissipation paths that all particles follow with a specific PRI size, WDM, link clock frequency etc. Within an iteration, a particle tracks the personal best solution (p_k), which is the best solution found by the particle k, and the global best solution (g_k), which is the best solution that was found by the entire population. Every particle moves towards the better solutions with some velocity and position. The computation step includes some amount of randomness instead of following an exact profile. This randomness can produce a superior solution which may result in other particle being attracted towards it. Each particle updates its velocity and position based on the following set of equations

$$\boldsymbol{v_{k+1} = wv_k + c_1 r_1 (p_k - x_k) + c_2 r_2 (g_k - x_k)} \quad (1)$$

$$\mathbf{x_{k+1} = x_k + v_{k+1}} \quad (\mathbf{2})$$

where, the current velocity and position for each particle is defined by v_k and x_k respectively, p_k designates current best solution based on particle k's history and g_k defines current best solution based on the entire population or swarm. The positive inertial weight w is assigned to control how fast or slow each particle can move based on its own weight or inertia, c_1 and c_2 are constant numbers denoted as learning parameters that control the learning rate of global vs. local optima, i.e., higher the weights, the faster the particles gravitate towards the current best solution. Instead of just following the current best solution in a linear path, r_1 and r_2 are random numbers from *0* to *1* that change every iteration, adding randomness to the path thus, finding newer better solutions on the way. The *stability* of the PSO algorithm is one of the key concerns where position and velocity can diverge instead of achieving convergence. To ensure solution convergence, we tune the learning and inertial weights carefully and also implement a velocity limit parameter V_{max}, where if the updated velocity exceeds the velocity limit, we saturate the velocity value to V_{max}.

6.4.2.2 ANT COLONY OPTIMIZATION (ACO)

The Ant Colony Optimization (ACO) metaheuristic was initially proposed by Colorni, Dorigo, and Maniezzo [167] with the fundamental idea inspired by the behavior of real ants, specifically, the way they organize efforts to collect food. ACO is a probabilistic technique for solving computational problems which can be reduced to finding good paths through graphs. In recent years, this algorithm has been applied to many combinatorial optimization problems such

as the asymmetric traveling salesman problem [51] and the graph coloring problem [52]. Although by itself, an ant is a simple and unsophisticated creature, collectively a colony of ants can perform useful tasks such as building nests, and foraging (searching for food). Ants achieve *stigmergic* communication by laying down a chemical substance called *pheromone* which can be sensed by other ants. When a pheromone trail laid by an ant that has found food is discovered by other ants, they tend to stop moving randomly and start following this specific trail, returning and reinforcing it if they eventually find food. Over time however, the pheromone trail starts to evaporate, thus reducing its attractive strength. The more time it takes for an ant to travel down the path and back again, the more time the pheromones have to evaporate. A short path, by comparison, gets marched over more frequently, and thus the pheromone density becomes higher on shorter paths than longer ones. Pheromone evaporation is crucial for *avoiding the convergence to a locally optimal solution*. If there were no evaporation at all, the paths chosen by the first ants would tend to be excessively attractive to the following ones. In that case, the exploration of the solution space would be constrained. The idea of the ACO based NoC synthesis algorithm is to mimic this behavior with "simulated ants" walking around a graph representing the problem to solve.

In our ACO implementation, the ant system represents source and destination core communications for the specific embedded application. Upper bounds for the power and latency for each edge in the core graph *G(V,E)* are computed using a baseline electrical NoC. As the flits are routed from the source to the destination, the ACO process can select among various values for the parameters in Table 6, such as the WDM density and PRI size. At the end of the route, power and latency for the selected configuration is compared against the upper bound. If the achieved results are better than the upper bound or the previous best result, then the pheromone

and likelihood values are updated. Gradually one dominant solution emerges. This best configuration satisfies latency constraints and has the lowest power dissipation paths that all ants follow with a specific PRI size, WDM, etc.

An ant in our formulation can be thought of as a simple computational agent. It iteratively constructs a solution for the problem at hand. The intermediate solutions are referred to as solution states. At each iteration of the algorithm, an ant *k* moves probabilistically from a state *i* to state *j*. Each of the parameters from Table 6 has a separate evaporation *trail value* (τ_{ij}) that represents the amount of pheromone deposited for a state transition between *i* and *j*. The selection probability for a parameter is a function of its attractiveness η_{ij}^{β}, defined by inverse of normalized power consumption for parameter β. Global convergence within the selection process is achieved by increasing attractiveness η_{ij}^{β} for low power dissipation solutions that meets latency constraints. An empirically- derived pheromone evaporation coefficient (ρ) with a value $1 > (\rho) > 0$ is utilized to control the evaporation of a trail over time. Trails are updated usually when all ants have completed their solution, increasing or decreasing the value of trails corresponding to moves that were part of "good" or "bad" solutions, respectively. $\Delta\tau_{ij}$ represents the change in trail value based on the choices available for a parameter, and the impact they have on the cost function (in our case power dissipation). At the start of simulation, selection probability of each parameter is equal. If power dissipation reduces significantly based on a parameter change for a majority of the communications, then $\Delta\tau_{ij}$ increases which causes the resulting selection probability to also increase. The selection for each parameter is performed using the following rules:

$$\boldsymbol{\tau_{ij}(t+n) = \rho * \tau_{ij}(t) + \Delta\, \tau_{ij}} \tag{3}$$

which is the trail update relation, with $\Delta\tau_{ij}$ given by:

$$\Delta\,\boldsymbol{\tau_{ij}} = \sum_{\mathbf{k=1}}^{\mathbf{m}} \Delta\,\boldsymbol{\tau_{ij}}^{\mathbf{k}} \tag{4}$$

for all *m* ants. The probability p_{ij}^{k} of moving from state *i* to *j* for the k^{th} ant is given as:

$$\mathbf{p_{ij}^{k}} = \frac{\boldsymbol{\tau_{ij}^{\alpha}}\,(\boldsymbol{t})\,\boldsymbol{\eta_{ij}^{\beta}}\,(\boldsymbol{t})}{\sum \boldsymbol{\tau_{ij}^{\alpha}}\,(\boldsymbol{t})\,\boldsymbol{\eta_{ij}^{\beta}}\,(\boldsymbol{t})} \tag{5}$$

This probability depends on the attractiveness η_{ij}^{β} of the move computed based on increasing a tunable weight for an ant for which power is lower and latency is within the constraints, and the trail level τ_{ij} of the move, indicating how proficient it has been in the past to make that particular move. $\alpha \geq 0$ is a parameter to control the influence of τ_{ij}, and $\beta \leq 1$ controls the influence of η_{ij}. Figure 51 shows the pseudo-code for our ACO formulation. The algorithm starts by calling *InitializeAntSystem()* to initialize the ant system, with each ant representing a communication trace. The function also sets up equal selection probability for every parameter. The function *UpdateAntSystem()* updates the probabilities of the individual ants, using relations *(3), (4),* and *(5)*. If the source core lies within a PRI region, the flow (ant) is directed towards the nearest gateway interface. The state transition parameter selection probability p_{ij}^{k} is applied to select serialization degree, clock frequency, flit width, and PRI data threshold. Once a flit reaches the uplink, number of waveguides and WDM density are selected for the next state. The same process is repeated for the destination gateway interface and destination core. As ants reach the destination, trail values are updated based on *(4)* and *(5)*, improving selection probability of

parameters that lead to lower power dissipation. At the end of the evaluation loop, the trail and pheromone updates are performed by calling *UpdateTrailMatrix()*. This process continues until a dominant solution emerges.

```
done = 1
while(done)
   // generate m number of ant systems; start with equal probability for each state
transition and update the system probability as we build the entire solution
InitializeAntSystem();
  for (j=0; j++; j < size(ant system))
    for (k=0; k++;k < linklength)
   // Choose the probability to move the flit from current state to next state and append the
chosen move to the k-th ant's set tabu_k until ant k has completed its solution
    UpdateAntSystem();
    end for

   ComputeNoCResults();    // trail update for each ant
   UpdateTrailMatrix();      // end of the loop, almost all ants will follow same trails

  end for

   // use cycle-accurate simulations to validate if latency constraints are satisfied
  if Termination criterion met then

     done = 0
  else done = 1 //Continue with next ACO iteration
  end if

end while
```

Figure 51 Ant colony optimization formulation

6.4.2.3 SIMULATED ANNEALING (SA)

Simulated Annealing (SA) algorithms [56] [163] [168] generate solutions to optimization problems using techniques inspired by annealing in solids. SA algorithms simulate the cooling of a metal in the heat bath known as annealing where the structural properties depend on the cooling rate. When a metal is hot and in liquid state, if cooled in a controlled fashion, large and consistent grains can be formed. On the other hand, grains can contain imperfections if the liquid is quenched or cooled rapidly. By slowly lowering the temperature, globally optimal solutions can be approached asymptotically. SA allows hill climbing or worse moves (with inferior quality) to be taken within the initial part of the iteration process. Based on the law of thermodynamics, at temperature t the probability of an increase in energy of magnitude δE is given by:

$$\mathbf{P(\delta E) = e^{\left(-\frac{\delta E}{kt}\right)}} \qquad (6)$$

where k is the Boltzmann's constant. This equation is directly applied to SA by dropping the Boltzmann constant which was only introduced into the equation to cope with different materials. The probability of accepting a state in SA is:

$$\mathbf{P = e^{\left(-\frac{c}{t}\right)} < r} \qquad (7)$$

where c defines the change in evaluation function output, t defines current temperature which is decremented at every iteration by some regression algorithm such as a linear method $t_{(k+1)}=\alpha.t_{(k)}$, with $\alpha < 1$, and r is a random number between 0 and 1.

An SA algorithm involves the evolution of an individual solution over a number of iterations, with a fitness value used for evaluating solution quality whose determination is problem dependent. At each iteration, individual parameters are selected randomly and the probability of accepting a solution is determined by equation *(7)*. A high enough starting temperature is selected to allow movement through the entire search space. As the algorithm progresses, the temperature is cooled down to confine solutions, allowing better solutions to be accepted until the final temperature is reached. As SAs are heuristics, the solution found is not always guaranteed to be the optimal solution. However in practice, SA has been used successfully to generate fairly high quality solutions in several problem domains.

```
done = 1
while(done)
    for (i=0; i++; i< N_ITER)
      GenerateInitialSolution()   // Initial solution
      ScheduleCoolingRate()   //Evaluate solution at cooling rate

      ComputeFitnessValue()  // Update fitness value
    end for

    if Termination criterion met then
       done = 0
    else
      done = 1 //Continue next N_ITER generations
    end if

end while
```

Figure 52 Simulated annealing algorithm formulation

Figure 52 shows the pseudo-code for our SA formulation of the synthesis problem. Our SA implementation begins with the calling of *GenerateInitialSolution()* to generate an initial solution. We utilize a genetic algorithm (GA) as an initial heuristic (explained in Section 5.2.4) to ensure high quality for the SA seed. Each GA chromosome consists of constituent parameters as defined in Table 6. At the end of 200 generations we use the best solution with the highest fitness value as a seed for SA. Subsequently, four key parameters for annealing are initialized by calling *ScheduleCoolingRate()*: *(i)* Starting temperature, *(ii)* Temperature decrement, *(iii)* Final temperature, and *(iv)* Iterations at each temperature. We tuned the starting temperature to be hot enough to allow our hybrid NoC parameters to traverse farther along in the solution space. Without this consideration the final solution would be very close to the starting SA solution. Based on the number of iterations for which the algorithm will be running, the temperature needs to be decremented such that it will eventually arrive at the stopping criterion. We also need to allow enough iterations at each temperature such that the system stabilizes at that temperature. We evaluated another method first suggested in [169] that proposes implementing one iteration at each temperature by decreasing the temperature very slowly. The formula we used was $t_{(k+1)} = t_{(k)}/(1+\beta t_{(k+1)})$ where β is a suitably small value as defined in [169]. However the approach did not yield any benefits in terms of improvement in results. As SA is a stochastic search algorithm, it is difficult to formally specify convergence criteria based on optimality. The results are expected to get better with every step, however sometimes the fitness of a solution, calculated by calling *ComputeFitnessValue()*, may remain unchanged for a number of cooling steps before any superior solution can be created.

```
done = 1
while(done)
  for (i=0; i++; i< N_ITER)
     // generate new system configuration based on initial population
     GenerateInitialPopulation()
     Crossover ()   //  current solutions are selected for mutation
     Mutation ()   //Evaluate the chromosome with upper bound

     ComputeNoCResults()
     ComputeFitnessValue()
  end for
  if Termination criterion met then

     done = 0
  else
     done = 1 //Continue next N_ITER generations
  end if

end while
```

Figure 53 Genetic algorithm formulation

6.4.2.4 GENETIC ALGORITHM (GA)

Genetic algorithms (GAs) [170] generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover. A GA involves the evolution of a population of individuals over a number of generations. Each individual of the population is assigned a fitness value whose determination is problem dependent. At each generation, individuals are selected for reproduction based on their fitness value. Such individuals are crossed to generate new individuals, and the new individuals are

mutated with some probability. The objective of a GA is to find the optimal solution to a problem. However, because GAs are heuristics, the solution found is not always guarantied to be the optimal solution. Nevertheless, experience in applying GAs to a variety of problems has shown that often the goodness of the solutions found by GAs is sufficiently high.

Figure 53 shows the pseudo-code for the GA formulation of our synthesis problem. Our GA implementation begins with the generation of an initial population by calling *GenerateInitialPopulation()*. Each individual element consists of a chromosome with constituent parameters as defined in Table 6. Based on empirical analysis, we set our GA population size to 2000, composed of chromosomes with parameter values set according to a uniform random distribution. The fitness value assigned to each chromosome consists of a weighted combination of average packet latency and communication power dissipation. The fitness is evaluated analytically based on the communication requirements of the application for which the hybrid NoC is being synthesized. Each application can have a unique set of communication patterns (represented by edges in the core graph), and thus the same architectural optimization (e.g., changing PRI size) can impact the latency and power dissipation of different applications differently. Similar to a roulette wheel, a probability based selection process was implemented for choosing chromosomes from the population, based on the relative fitness value. Crossover was applied to randomly paired parameters by exchanging genetic information via swapping bits within the parent's chromosome calling *Crossover()* We also implemented multipoint crossovers where multiple parts of chromosome strings replaced each other. Then mutation was performed by calling *Mutation(),* where one parameter was changed within allowable limits (Table 6). Mutations and crossovers produced the next generations. Individuals with the crossover and

mutations generate new offsprings that replace original chromosomes if the offsprings satisfy upper bound with *ComputeNoCResults()*.

Since GA is a stochastic search algorithm, it is difficult to formally specify convergence criteria based on optimality. The results are expected to get better with every generation, however sometimes the fitness of a population, calculated by calling *ComputeFitnessValue()*, may remain unchanged for a number of generations before any superior chromosomes can be created. The general practice is to terminate the GA after a predefined number of generations and then to evaluate the quality of the results within the population against the expected optimal where expected optimal is obtained using extended GA runs or iterations.

6.5 CYCLE ACCURATE SIMULATION AND VALIDATION

Upon completion of the synthesis algorithms, we verified our results based on cycle accurate SystemC [158] simulator. If power and performance of synthesized simulation does not match within 5% of the cycle accurate simulator, we repeated the synthesis process until we correlate results. This is achieved by ultimately calling *ValidateResults()* to validate the best solution by using our in-house SystemC-based [158] cycle-accurate hybrid nanophotonic-electric NoC simulator. This is done to ensure that latency constraints are satisfied in the presence of communication congestion and computation delays, which can only be accurately analyzed via simulation

6.6 EXPERIMENTS

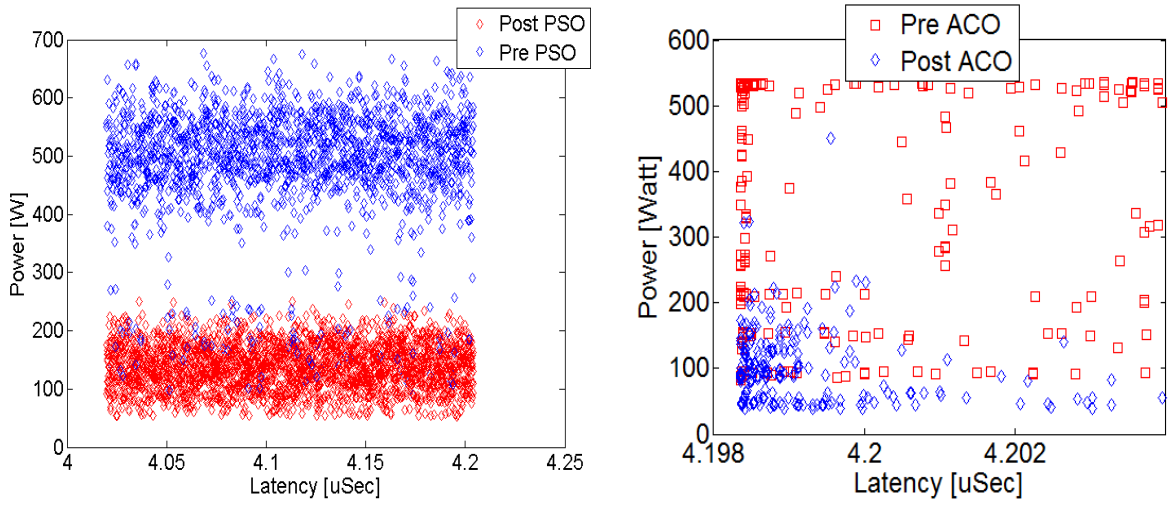
6.6.1 EXPERIMENTAL SETUP

We conducted experimental analysis to compare the performance of our PSO, ACO, GA and SA based synthesis frameworks for mid-size *6×6 (36-core)* and large-size *10×10 (100-core)* CMPs with a 2D mesh hybrid photonic ring/mesh NoC fabric. Parallel implementations of seven *SPLASH-2* benchmarks [135] (*barnes, lu, cholesky, fft, fmm, radiosity, radix*) were utilized to guide the application-specific synthesis. We also implemented *NAS* [136] and *PARSEC* [137] parallel application benchmarks. *NAS* [136] benchmarks are derived from computational fluid dynamics (CFD) applications. The Princeton Application Repository for Shared-Memory Computers (*PARSEC*) [137] suite is composed of several multithreaded programs that represent next-generation shared-memory programs for CMPs. Our synthesis runs lasted around *8* to *10* hours for each search algorithm; however initial runs lasted around *4-6* days. Once We realized that 8-10 hours of runtime was able to provide solutions within *2-4*% of solutions generated with extended runs, We reduced our simulation time to be more efficient.

We targeted a *32nm* process technology with the assumption of a 400 mm^2 die area budget. shows delay values for *32nm* technology that we assumed, obtained from [138] and from device fabrication results [140]. The delay of an optimally repeated and sized copper wire at *32nm* was assumed to be *42ps/mm* [29]. The power dissipated in the hybrid photonic NoC can be categorized into *(i)* electrical network power and *(ii)* photonic ring network power. The static and dynamic power dissipation of electrical routers and links was derived from Orion 2.0 [141]. For the energy dissipation of the modulator driver and TIA power we used ITRS device projections [4] and standard circuit procedures. An off-chip electrical laser power of 3.3W (with 30% efficiency) is also considered in our energy calculations. The laser power value accounts for per

component optical losses due to non-linearity (1dB at 30mW), couplers/splitters (*1.2dB*), waveguides (*3dB/cm*), waveguide crossings (*0.05dB*), ring modulators (*1dB*), receiver filters (*1.5dB*) and photodetectors (*0.1 dB*).

The search heuristics were configured as follows. For the PSO algorithm, we empirically set the inertia weight $w = 0.66$, $c1=c2=0.5$, and the velocity limit parameter $V_{max} = 0.33$. For the ACO algorithm, we set the pheromone evaporation coefficient $\rho = 0.67$, and tunable weights α and β were set to *0.46* and *0.54*, respectively. For the SA algorithm, we set α to *0.997*, and utilized initial temperature $T_0 = 1000^o\ C$. For the GA, we maintained an initial population size of $M = 256(N \times N)^2$ where N is the (X or Y) mesh dimension and ran the algorithm for up to *2000* generations. We evaluated our GA implementation for various mutation and crossover probabilities and ultimately utilized values of *0.3* and *0.2* respectively.



(a) **(b)**

Figure 54 Pre and post PSO and ACO power consumption and latency for solution space of lu benchmark

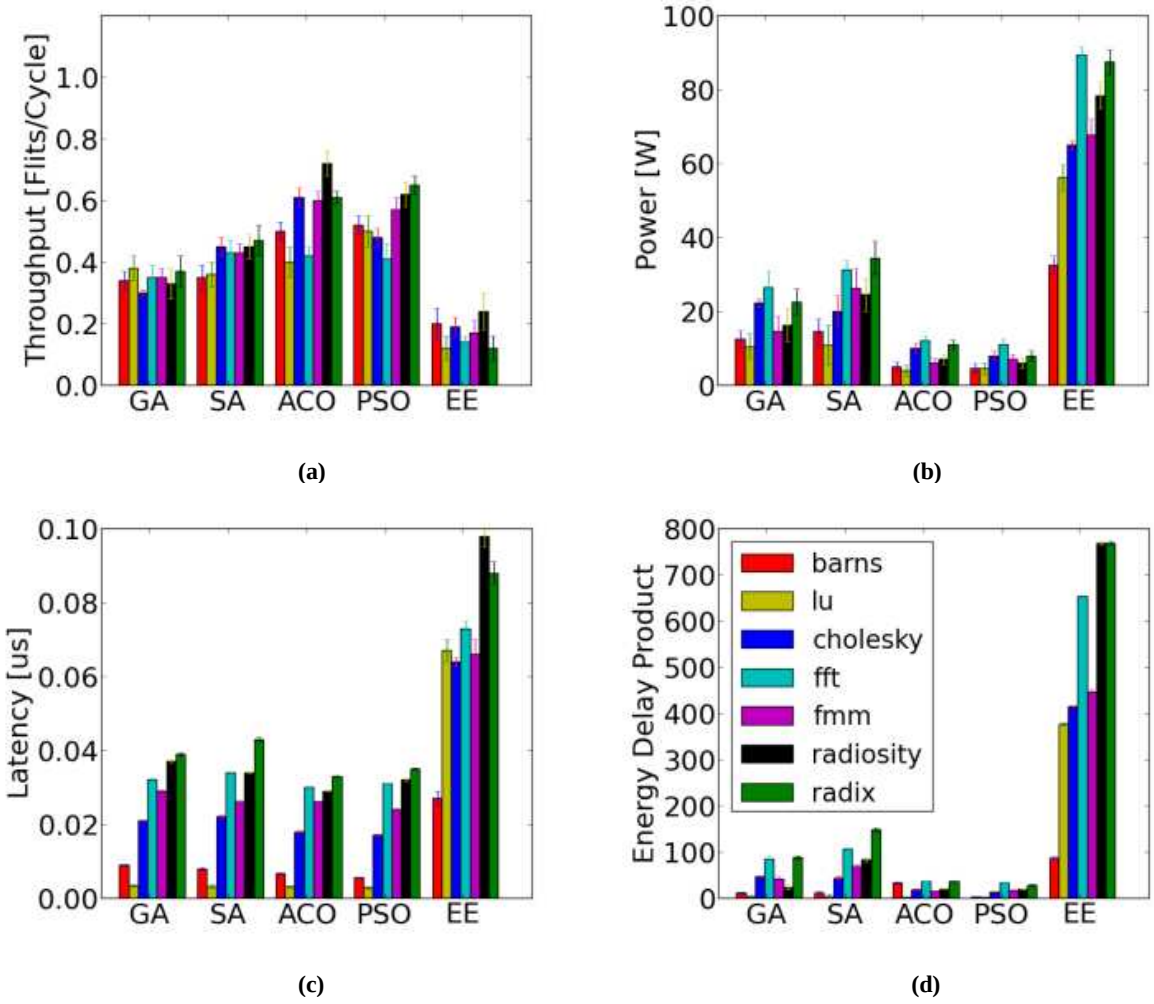


Figure 55 Latency, throughput, power, and energy-delay product comparison for *SPLASH-2* benchmarks for 10x10 NoC

6.6.2 RESULTS

Our first experiment provides insights into the workings of the PSO and ACO algorithms. Figure 54 shows the solution space pre- and post-PSO and ACO, that compares the power and average packet latency, for the *lu* benchmark from the SPLASH-2 suite. The solution space is relatively randomly distributed in 2D with higher power consumption before the PSO and ACO algorithms begin execution. The ACO solutions swarm towards the shortest path lower end of the

2D space observed in Figure 54 *(b)*. This result can be explained based on equation (*5)*, as attractiveness of shortest paths grow higher within ACO algorithm. On the other hand, PSO algorithm drives the solutions towards lower power per Figure 54 *(a)* by following velocity/position profiles relative to local and global minimum power solutions per equation (*1)*. This indicates an improvement in power while maintaining average packet latency characteristics, and shows that the ACO approach leads to desirable quality solutions.

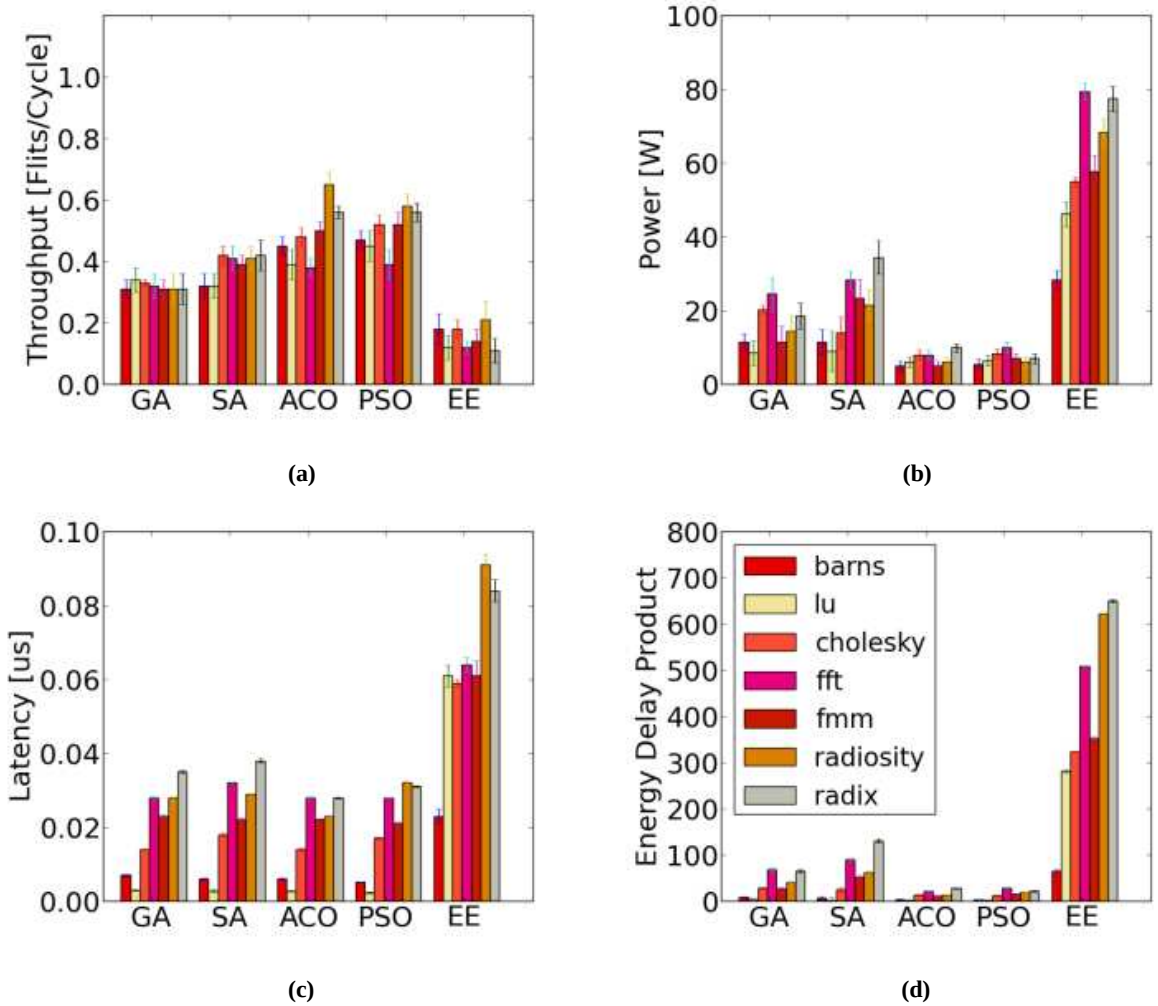


Figure 56 Latency, throughput, power, and energy-delay product comparison for *SPLASH-2* benchmarks for 6x6 NoC

To gain further insight regarding the quality of the generated results, we evaluated various parameters of the best solutions generated by ASO, PSO, GA and SA for the seven SPLASH-2 benchmarks. Figure 55 shows results for the 10x10 CMP while Figure 56 presents results on a 6×6 CMP implementation. For benchmarks that require more frequent local and global communication, ACO utilizes a higher degree of photonic path communications while enabling higher clock frequency for the electrical path and achieving an elegant trade off. Also to enable higher photonic path communication, the PSO and ACO algorithms generate solutions with greater number of uplinks. The latency and WDM degree results indicate that PSO and SA each have a unique set of benchmarks for which the synthesized architecture provides lower average packet latency and lower WDM than the solution generated by the other approach. As far as the number of uplinks are concerned, with the exception of *radix* and *radiosity*, PSO and SA both select the same number of links for their best solutions. Table 7 - Table 10 summarize the 10x10 hybrid photonic NoC solutions generated by PSO and ACO algorithms respectively for the SPLASH-2 benchmark applications. The communication traffic pattern for each application is different, so these results provide insights into the inner workings of each synthesis approach. As the communication traffic goes up, both algorithms tend to adapt differently towards solution configurations. Runtime configuration can enable custom solutions that balance various trade-offs, for example PSO and ACO adapts more efficiently to higher PRI size for *radix* and *radiosity* than SA and GA as shown in Table 9 and Table 10. Both algorithms successfully increase WDM degrees as core to core communication increases. The PRI data threshold M_{th}, diverts communication through photonic channels if data length exceeds beyond this limit. The PSO algorithm optimizes the M_{th} limit to a lower number than SA thus increasing the volume of data traversing the photonic communication path. We also monitored number of average hops from the

source and destination cores to the uplinks within a PRI region to better understand how far data packets needed to travel to reach an uplink. The number of PSO generated hops was higher than ACO confirming the consistency with lower PRI threshold achieved by PSO, particularly for the *radiosity*, *fmm* and *radix* benchmarks.

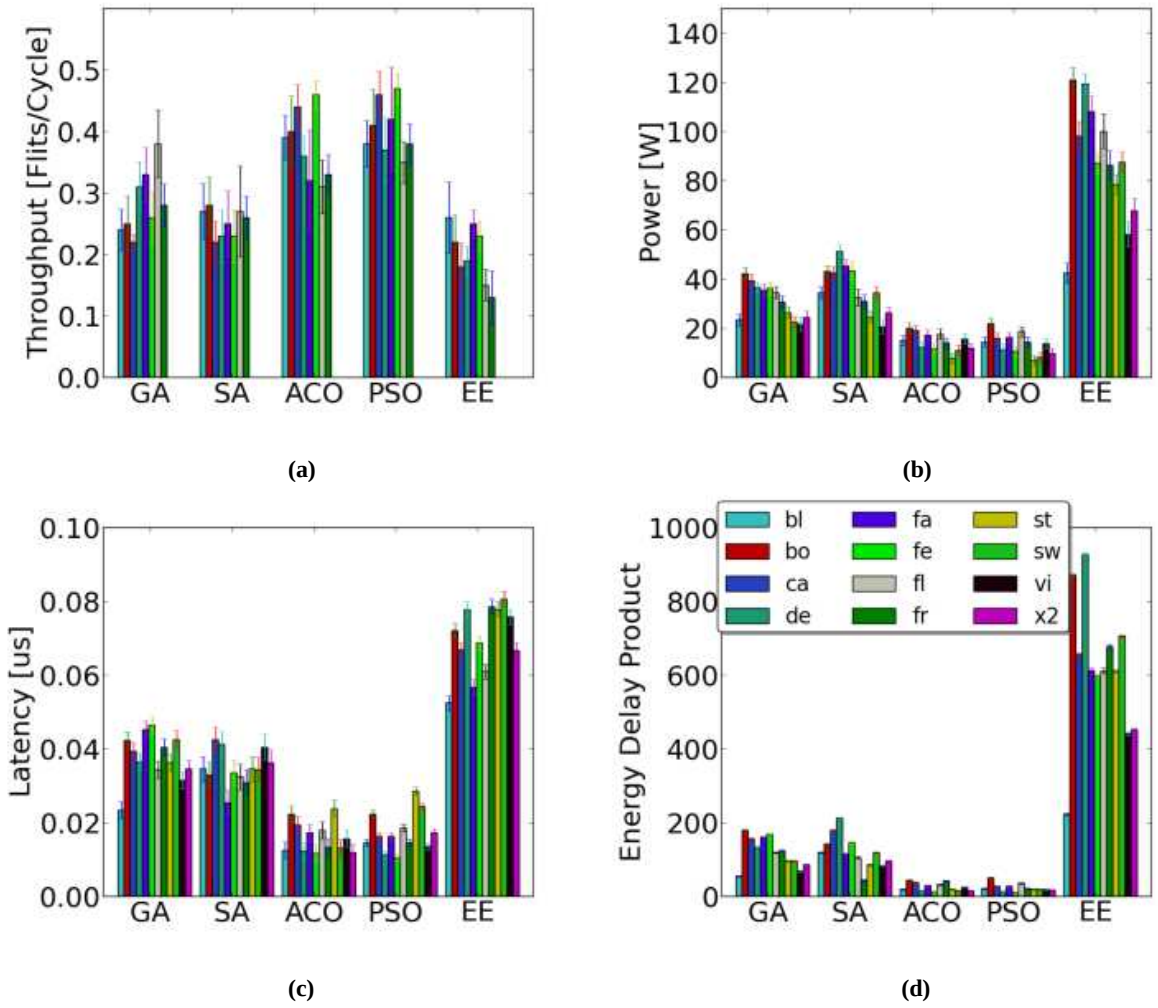


Figure 57 Latency, throughput, power, and energy-delay product comparison for *PARSEC* benchmarks for 10x10 NoC

Despite the overhead of a separate photonic layer, it can be seen that using hybrid photonic NoCs can lead to significant orders of magnitude savings in power dissipation. Among the synthesized approaches, it can be seen that the PSO and ACO generates solutions that are more power efficient than those generated by the SA and GA. The synthesized solutions with PSO and ACO have as much as 1.2× lower power dissipation than average solutions generated by SA and GA with up to 2.2× for the best case. Figure 57 and Figure 58 presents the best solutions generated by ASO, PSO, GA and SA for PARSEC benchmarks, and for the 10x10 and 6×6 CMP implementations respectively. Figure 59 and Figure 60 presents NAS benchmarks for similar configurations. Both benchmarks shows capability of our synthesis process achieving excellent improvements. Figure 61 summarizes the energy-delay product improvements for PSO, ACO over the GA and SA algorithms. We also observed significant (up to 18×) improvements with PSO and ACO generated hybrid photonic NoC solutions compared to the baseline 2D electrical mesh NoC architecture. Our novel implementation of PSO achieves average 64% energy-delay improvements over GA and 53% over SA while the ACO implementation achieves 107% energy-delay improvements over GA and 62% over SA.

Table 7 PSO synthesis results

Synthesis Parameters	*Barns*	*Lu*	*Cholesky*	*Fft*	*Fmm*	*Radiosity*	*Radix*
WDM	68	122	44	83	135	132	143
Uplinks	4	4	4	4	4	8	12
PRI	15	12	10	10	12	12	12
PRI Data Threshold	96	4	120	48	7	54	96
Clock Frequency	5	4	5	4	4	2	2
Source PRI Uplink	9	7	8	9	8	9	9
Dest PRI Uplinks	9	9	9	9	9	9	9
Flit Width	43	256	28	85	256	128	128
Serialization	6	1	9	3	1	2	2
Waveguides	18	12	135	136	112	20	18

Table 8 ACO synthesis results

Synthesis Parameters	*Barnes*	*Lu*	*Cholesky*	*Fft*	*Fmm*	*Radiosity*	*Radix*
WDM	128	56	145	138	56	67	166
Uplinks	4	4	4	4	4	8	12
PRI size	4	4	4	4	8	8	8
PRI Threshold	72	32	56	46	186	459	148
Clock Freq	4	2	4	5	4	5	5
Src Uplinks	4	4	4	4	4	8	12
Dest Uplinks	4	4	4	4	4	58	12
Flit Width	28	64	46	49	28	28	42
Serialization	12	4	3	8	12	18	10
Waveguides	48	32	58	108	128	128	256

Table 9 SA synthesis results

Synthesis Parameters	*Barns*	*Lu*	*Cholesky*	*Fft*	*Fmm*	*Radiosity*	*Radix*
WDM	102	79	58	128	93	141	63
Uplinks	4	4	4	4	4	4	4
PRI	4	4	4	4	4	4	4
PRI Data Threshold	176	16	60	96	280	459	168
Clock Frequency	4	2	4	5	6	5	4
Source PRI Uplink	4	4	4	4	5	5	4
Dest PRI Uplinks	5	5	4	4	5	5	4
Flit Width	23	64	85	43	26	15	37
Serialization	11	4	3	6	10	17	7
Waveguides	48	2	25	200	105	128	90

Table 10 GA synthesis results

Synthesis Parameters	*barns*	*Lu*	*Cholesky*	*Fft*	*Fmm*	*Radiosity*	*Radix*
WDM	89	73	56	89	67	130	67
Uplinks	4	4	4	4	4	4	4
PRI	4	4	4	4	4	4	4
PRI Data Threshold	223	23	45	78	139	320	123
Clock Frequency	3	3	3	3	4	4	3
Source PRI Uplink	4	4	4	4	4	4	4
Dest PRI Uplinks	4	4	4	4	4	4	4
Flit Width	28	78	56	43	22	12	38
Serialization	11	12	13	11	11	12	11
Waveguides	35	25	45	178	99	111	89

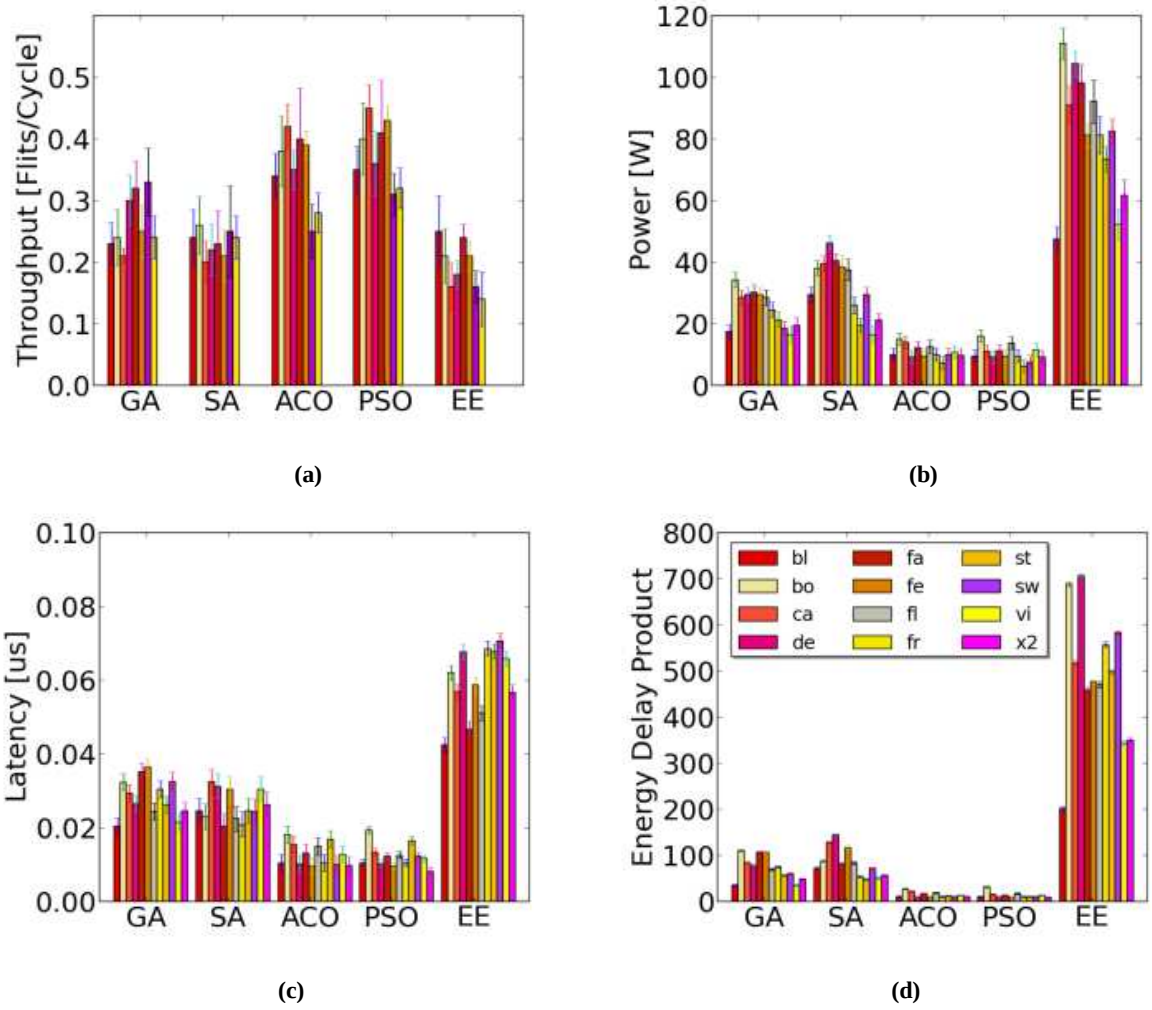


Figure 58 Latency, throughput, power, and energy-delay product comparison for *PARSEC* [40] benchmarks for 6x6 NoC

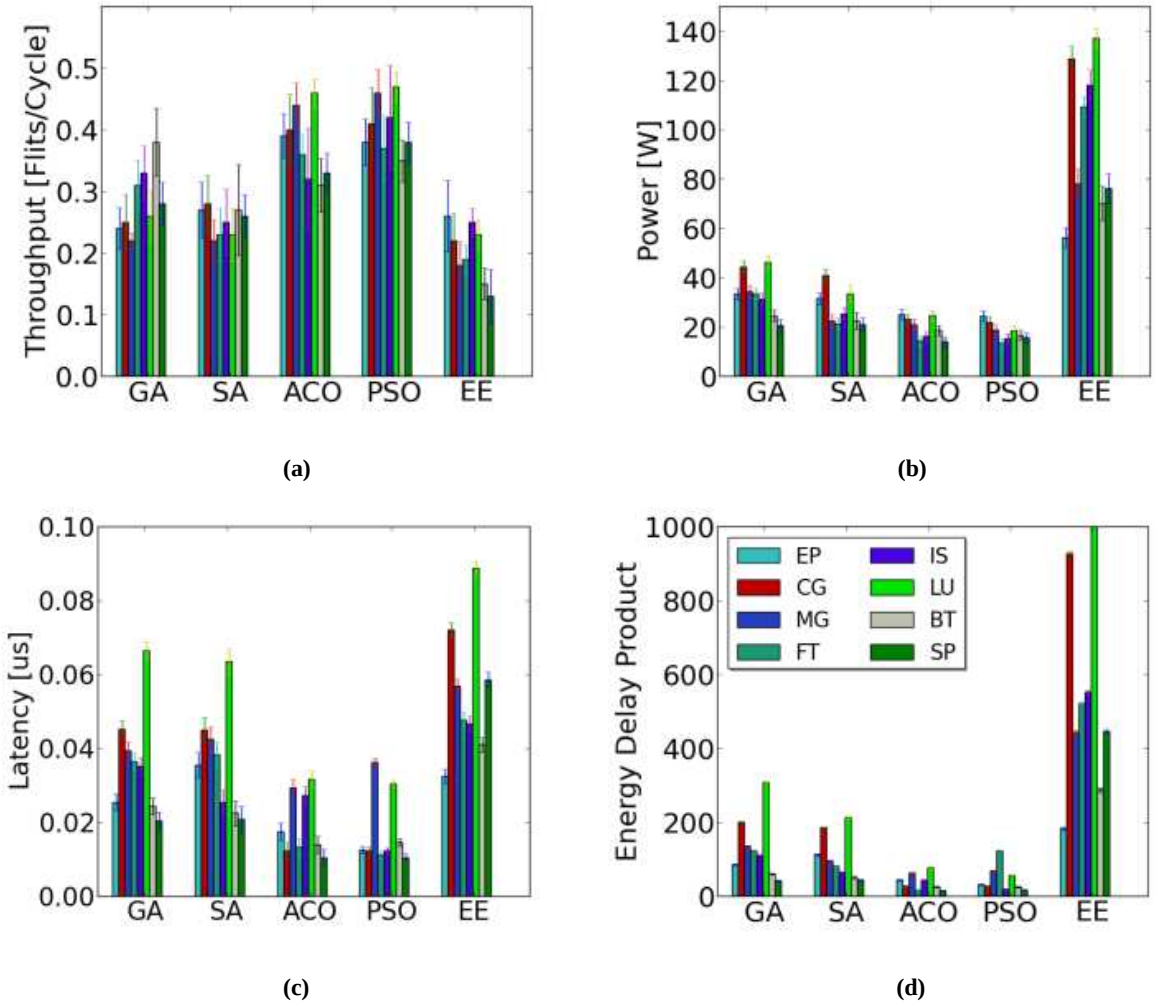


Figure 59 Normalized latency, throughput, power, and energy-delay product comparison *NAS* [39] benchmarks for 10x10 NoC

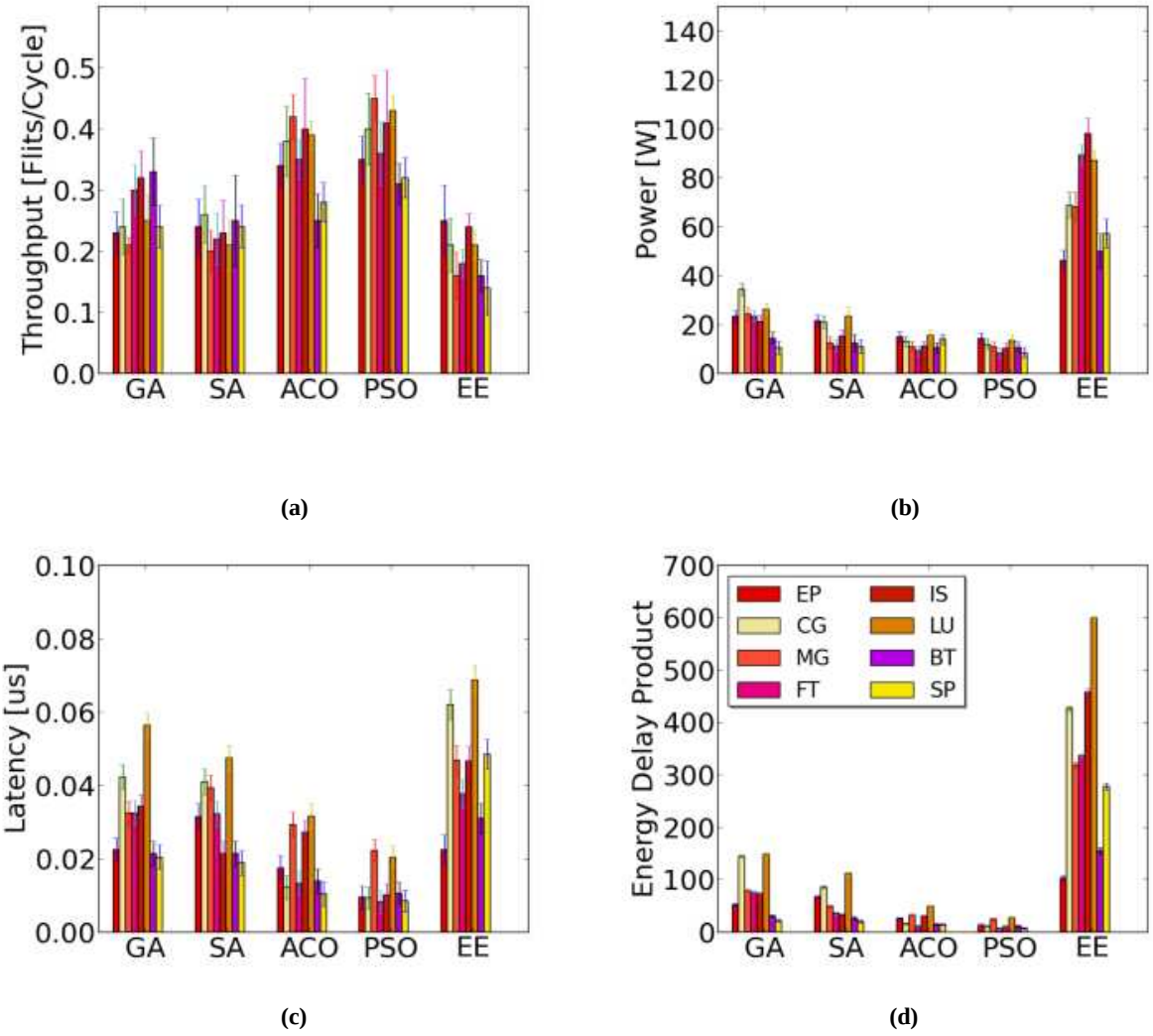
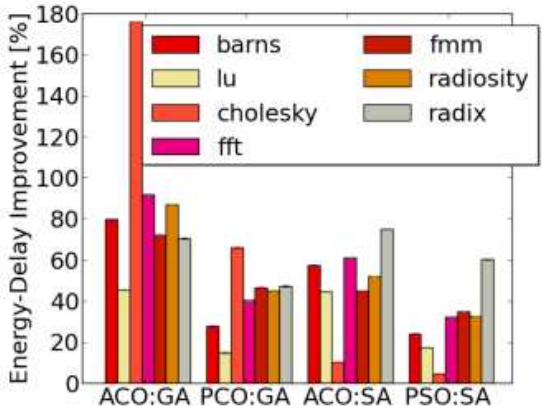
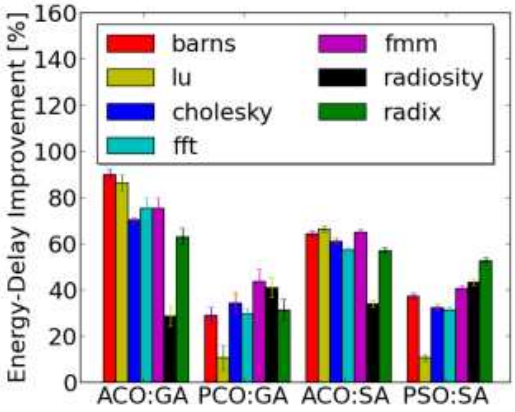


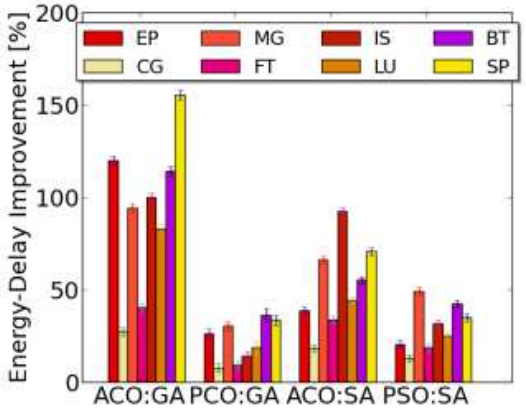
Figure 60 Latency, throughput, power, and energy-delay product comparison *NAS* [39] benchmarks for 6x6 NoC



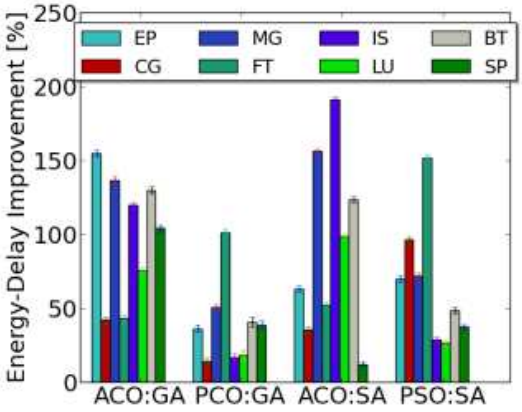
(a)



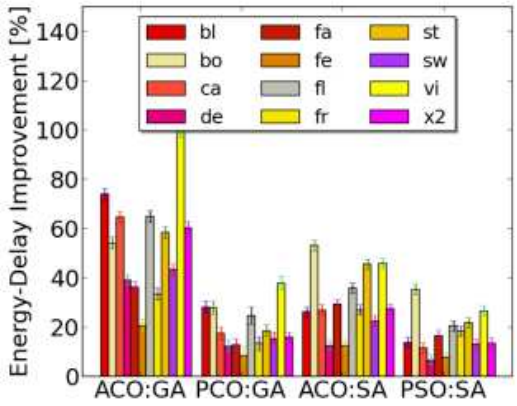
(b)



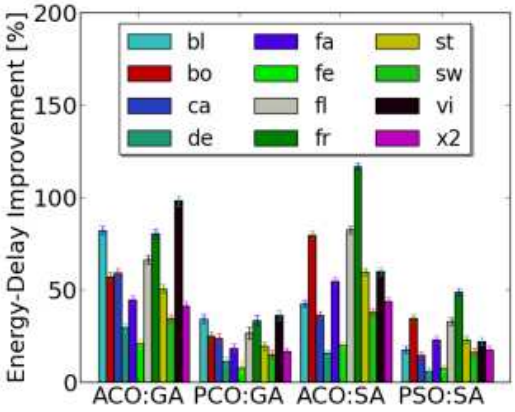
(c)



(d)



(e)



(f)

Figure 61 Energy delay product improvements for solutions generated by ACO and PSO over SA and GA for (a) 6x6 *SPLASH-2* (b) 10x10 *SPLASH-2* , (c) 6x6 *NAS*, (d) 10x10 *NAS*, (e) 6x6 *PARSEC*, (f) 10x10 *PARSEC* benchmarks

6.7 RESULT SUMMARY

In this chapter, we proposed a framework for synthesizing hybrid photonic NoC architectures for emerging CMPs. We formulate the synthesis problem using four different search heuristics: Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Simulated Annealing (SA) and Genetic Algorithms (GA). Our results and experimental data demonstrate significant promise for the ACO as well as PSO-based search heuristics for our problem domain of hybrid photonic NoC synthesis, allowing us to determine application-specific architectural parameters that minimize power dissipation while satisfying application latency constraints.

7 *HELIX:* DESIGN AND SYNTHESIS OF HYBRID FREE SPACE APPLICATION-SPECIFIC NOC ARCHITECTURES

Hybrid NoCs with nanophotonic guided waveguides and silicon microring resonator modulators impose many challenges such as high thermal tune up power, crossing losses, and high power dissipation. Due to these challenges productization of such architectures has yet to become commercially viable. Unfortunately, increasing embedded application complexity, hardware dependencies, and performance variability makes optimizing hybrid NoCs a daunting task because of the need to traverse a massive design space. *To date, prior work on automated NoC synthesis has mainly focused on electrical NoCs. For the first time, we propose a suite of techniques for effectively synthesizing hybrid photonic on-chip interconnects.* No prior work has addressed the problem of synthesizing application-specific hybrid nanophotonic-electric NoCs with an irregular topology to the best of our knowledge. Considering the above unaddressed major challenges, in this chapter we propose and discuss the *HELIX* framework for application-specific synthesis of hybrid NoC architectures that combine electrical NoCs with free-space nanophotonic NoCs. Based on our experimental studies, we demonstrate that the presented algorithms in this chapter produce superior NoC architectures when compared to algorithms proposed in prior work for electrical NoCs.

7.1 HYBRID PHOTONIC FREE SPACE NOC ARCHITECTURE OVERVIEW

To maximize performance in SoCs, ideally any two connected cores should communicate with each other using a point-to-point single hop network. For an $(m \times n)$ core SoC architecture, a single hop connectivity NoC fabric requires $O(m \times n)^2$ links. This is prohibitive to implement using a reasonable number of metal and photonic waveguide layers. However a free-space

optical interconnect (FSOI) network can eliminate much of the complexity of laying out multiple waveguides and also reduce global metal interconnect counts, while enabling 1-hop or 2-hop communication paths.

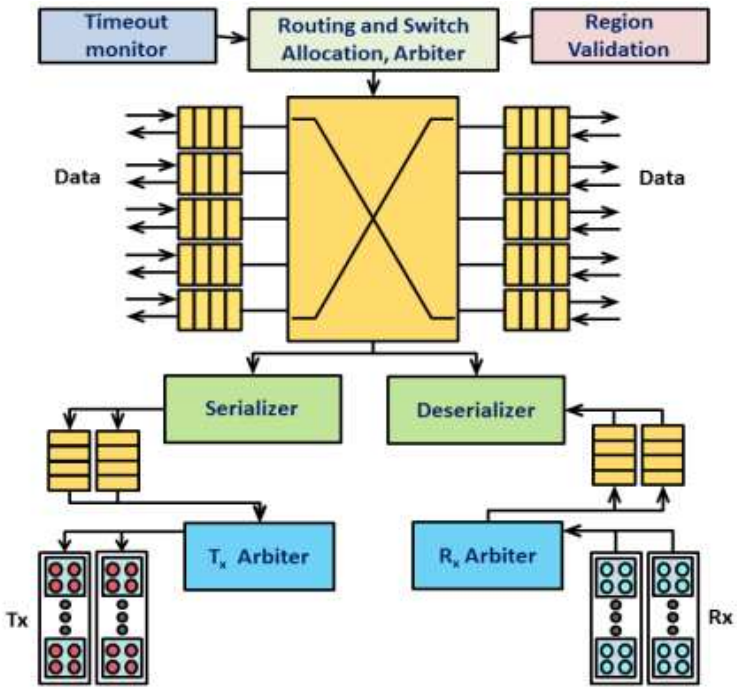


Figure 62 Gateway interface router architecture

Our chosen FSOI network fabric utilizes micro-mirrors and reflectors, with light traversing through free-space to achieve 1-hop or 2-hop transfers with low overhead. For a 1-hop *( )* SoC with flit width of *k*, each node needs 2× Gbps/(GHz CPU clock) MQW devices, while a 2-hop *( )* SoC with the same flit width needs 4× Gbps/(CPU Clock)) MQW devices [43]. As an example, for a 12×12 core SoC with a 1-hop NoC, with flit width of 256 bits at 40 Gbps/link and a 3.88 GHz CPU clock, *7322* MQW devices are required. The photonic components for a *20mm×20mm* SoC die size will consume < *5* mm^2 on-chip area for a 1-hop FSOI-based NoC with *100μm* MQW devices. In contrast, a 2-hop

NoC will require only *1128* MQW devices with < 1 mm^2 area and a 5.5× power reduction over a 1-hop NoC, but at the cost of system bandwidth drop from *300 to 45* Tbps. We explore hop-count selection on a per-communication flow basis to enable power-bandwidth trade-offs in our framework.

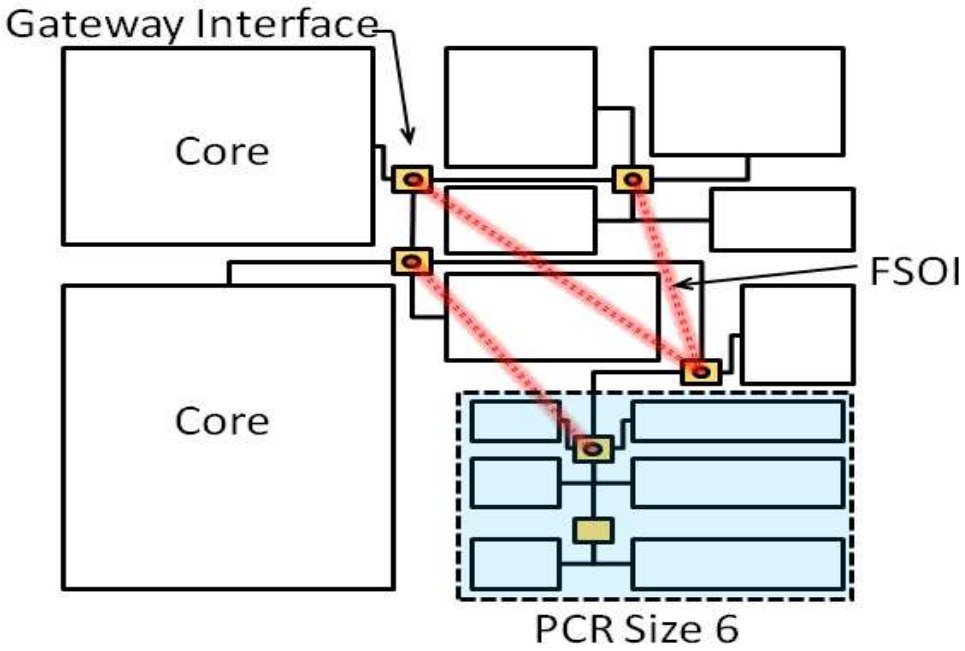


Figure 63 Photonic concentration region (PCR)

We consider a SoC platform with a dedicated photonic layer that supports FSOI links, interfacing with an electrical NoC. Our electrical NoC is composed of two types of routers: *(i)* conventional four stage pipelined electrical routers that have *n* I/O ports and interface with local cores; and *(ii)* gateway interface routers (Figure 62) that are also four-stage pipelined but have additional photonic ports (a total of *n+2* I/O ports). The photonic link interface in gateway routers is responsible for sending/ receiving flits to/from photonic links in the photonic layer. Both types of routers have an input and output queued crossbar with a 4-flit buffer on each

input/output port, with the exception of the photonic ports in gateway interface routers that use double buffering to cope more effectively with the higher photonic path throughput. The serializer/deserializer modules can support serialization degree of 2, 4, 8 and 16, and allow for very high optical I/O pad density. The resulting high-density 2D array of surface-normal optoelectronic MQW devices can provide the necessary intra chip bandwidth density without the complexity of wavelength division multiplexing (WDM) [171].

A unique feature of our hybrid NoC fabric is the reconfigurable traffic partitioning between electrical and photonic links. To minimize implementation cost, our synthesis framework limits the number of gateway interfaces. An adaptive photonic concentration region (PCR) ensures appropriate scaling and utilization with changing communication demands. A PCR is defined as the number of cores around the gateway interface that can utilize the FSOI path for communication (Figure 63). Cores within the same PCR communicate with each other via the electrical NoC (intra-PCR transfers). Cores that need to communicate and reside in different PCRs communicate using photonic paths (inter-PCR transfers). The electrical NoC transfers use XY routing, and a modified PCR-aware routing scheme for selective data transmission through the photonic links, with timeout-based regressive deadlock handling, based on the approach presented in [55].

Our arbitration approach is different from FSOI-based gateway interface routers proposed in [42] and [43] that utilize transfers without any arbitration. These routing schemes directly stream data to destination cores and manage collision of photonic data with a collision handling scheme (e.g., when multiple source nodes send data to the same destination core). But we observed that the performance benefits of eliminating arbitration are overshadowed by high penalties of collision handling and retransmission for high performance communication flows.

Therefore in our gateway interface routers we implemented support for reservation channels to reserve FSOI data paths. An additional input and output reservation channel port is added to the routers for this purpose.

7.2 SYNTHESIS PROBLEM FORMULATION

This section summarizes the inputs to our problem and formalizes our problem objective:

7.2.1 APPLICATION WORKLOAD CONSTRAINTS

- Application *communication trace graph G(V,M,L)* for each application in a multi-application workload, where $v_i \in V$ is a set of processing cores, $m_i \in M$ a set of memory blocks, $l_i \in L$ a set of directed communication links;
- Application-specific communication bandwidth constraints $\omega_{i,j}$ in bits/cycle and latency constraints $\lambda_{i,j}$ in cycles between $\{v_i, v_j\}$ or $\{m_i, m_j\}$;

7.2.2 SOC PLATFORM CONSTRAINTS

- X_{max} and Y_{max} are the maximum dimensions of the die along the X and Y axes; and the aspect ratio X_{die}/Y_{die} of the synthesized die should be between 0.9 – 1.1 to obtain an approximately square die layout;
- Each network link is constrained by a maximum length γ that represents the maximum distance a signal can travel in a single cycle, based on CMOS process technology;

7.2.3 PROBLEM OBJECTIVE

- Synthesize a hybrid nanophotonic-electric application-specific NoC architecture *J(R, L_e, L_p, C)* where *R* is a set of hybrid routers, L_e and L_p represents the set of electrical and photonic links, and *C* is a core-to-die mapping function; such that communication power is minimized while meeting bandwidth and latency constraints of the given application(s), and platform constraints of the SoC;

7.2.4 CONFIGURATION PARAMETERS

- Application task to core mapping;
- Layout of cores and memories on the planar die;
- Number and layout of hybrid electro-photonic and electrical-only routers that utilize a set of photonic $p_i \in P$ or electrical links $e_k \in E$ to support communication for a given multi-application workload;
- Size of photonic concentration region (PCR) that determines the cores/memories allowed to use each hybrid photonic router on the die (see Section 4 for details);
- Serialization degree D_n at electro-photonic interfaces;
- Hop count (1-hop or 2-hop) selection for FSOI links;

7.3 *HELIX* SYNTHESIS FRAMEWORK OVERVIEW

In this section, we present our novel framework for synthesizing hybrid nanophotonic-electric NoCs, which consists of the following steps as shown in Figure 64 *(i)* task-to-core mapping; *(ii)* floorplanning; *(iii)* Steiner tree based network formation; *(iv)* link clustering and dual level router mapping *(v)* PCR allocation *(vi)* conflict analysis and resolution; and *(vii)*

validation with cycle-accurate simulation. Due to lack of space, here we briefly discuss each step. The following subsections provide an overview of these steps.

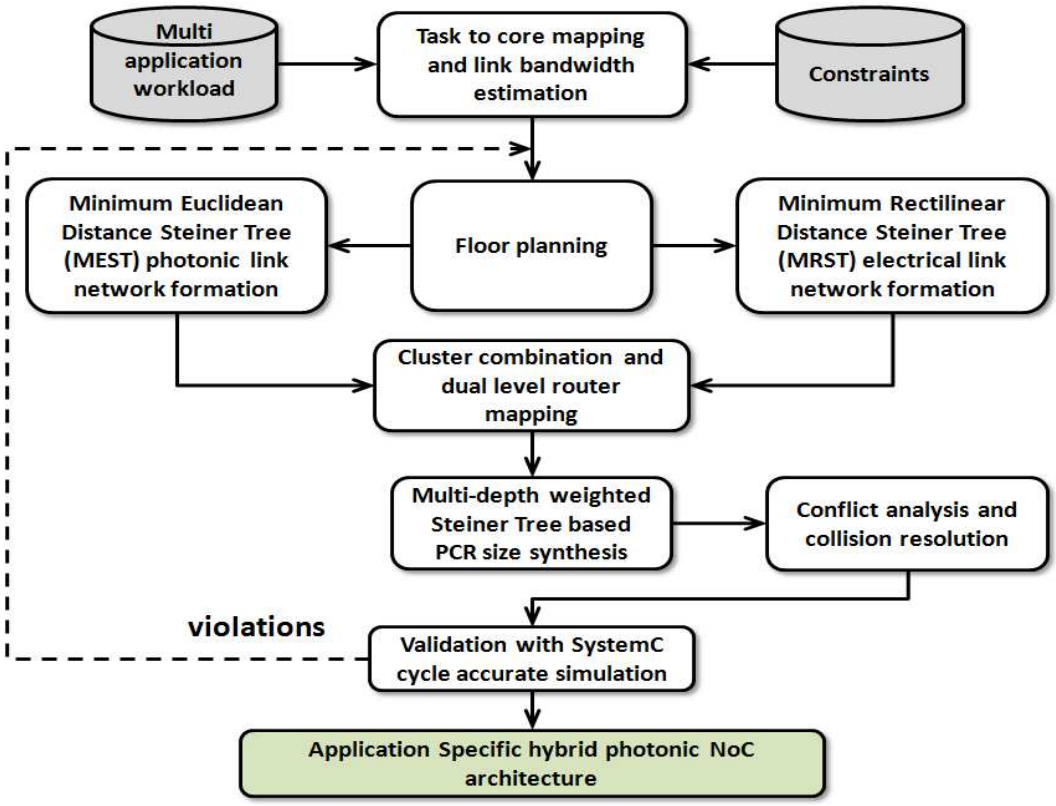


Figure 64 *HELIX* hybrid electro-photonic NoC synthesis flow

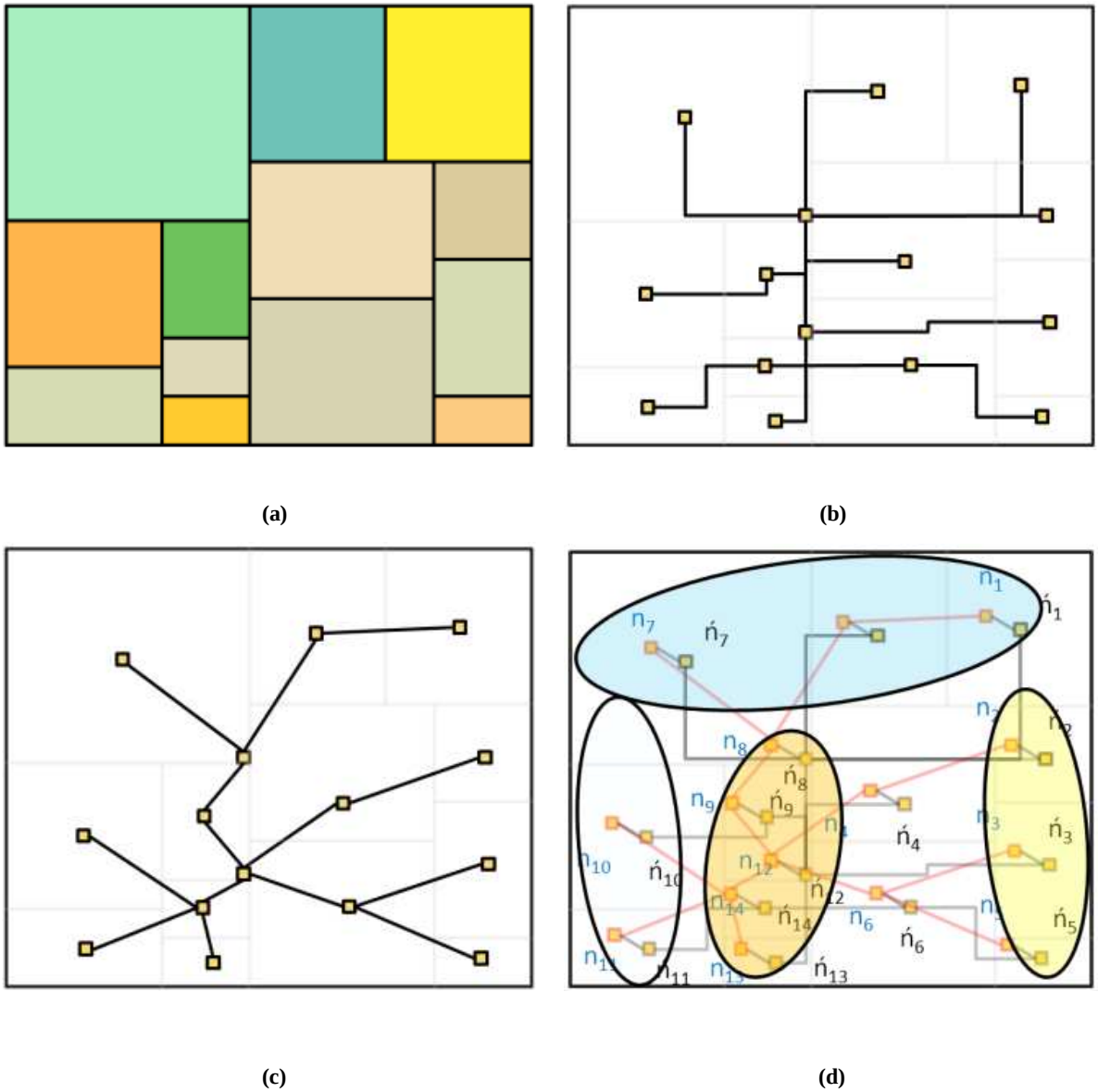


Figure 65 (a) Output of floorplanner (b) Minimum Euclidean Distance Steiner Tree (MEST) for electrical network (c) Minimum Rectilinear Distance Steiner Tree (MRST) for FSOI links (d) clustering and dual level router mapping

7.3.1 TASK TO CORE MAPPING

In this first step, we perform task-to-core mapping and link bandwidth estimation. The step involves mapping of n tasks to m heterogeneous cores for the given application(s) task flow graph. We perform task execution-time estimation as well as estimation of inter-core data transfers using an instruction simulator [172]. We implement a genetic algorithm (GA) [170] to

accomplish task to core mapping. The GA chromosome consists of possible mappings for the given application tasks to available cores, as well as a virtual link (electrical or FSOI) between cores that can satisfy bandwidth and latency constraints at a coarse granularity. The specific parameters used in the GA implementation are described in detail in the experimental setup section. The GA cost function represents overall communication power and the GA attempts to create a task-to-core mapping to minimize this power.

7.3.2 FLOORPLANNING

For hybrid nanophotonic-electric NoC architectures, the floorplanning step is significantly more complex than traditional floorplanning, as the power consumption and delay of electrical wire and FSOI links differ significantly. To the best of our knowledge there is no floorplanning tool available that can support such hybrid FSOI and electrical wire based architectures. We therefore designed an enhanced system-level NoC floorplanning tool that uses mixed integer linear programming (MILP) to perform communication-aware and power-aware core placement on the die. Our MILP minimization objective function is a linear combination of the weighed power-latency and the overall chip area, representing the metrics that are optimized in this stage:

$$\left[\sum_{\forall\, \boldsymbol{c(u,v)}\in \boldsymbol{E}}^{\boldsymbol{i,j}} \boldsymbol{l(u,v)} \times \boldsymbol{Cp_{i,j}}\right] \propto +[\boldsymbol{X_{max}} + \boldsymbol{Y_{max}}]\boldsymbol{\beta} \tag{8}$$

where, $l(u,v) \times Cp_{i,j}$ is weighed communication power and link distance between cores, $\propto$ and β are constants, and X_{max} and Y_{max} are the maximum allowed dimensions of the die along the X and Y axes. The floorplanner also integrates an aspect ratio constraint, to achieve an approximately square shaped floorplan. As FSOI links consume less power than electrical links, the floorplanner allows placing cores communicating via FSOI links farther apart than cores communicating via electrical links. Note also that at this stage, routers have yet to be allocated,

and are assigned arbitrary locations with respect to cores (1 virtual router/core). The output of this step is a floorplan as shown in Figure 65 (a).

7.3.3 MEST AND MRST BASED NETWORK FORMATION

In this step, we generate Minimum Rectilinear distance Steiner Trees (MRST) for the electrical network and Minimum Euclidian distance Steiner Trees (MEST) for the free space photonic network. We implemented these separate tree structures because electrical signal transmission occurs through rectilinear wires, and their Manhattan distance is best captured by an MRST; and free space photonic transfers can occur using non-rectilinear links, and their Euclidian distances are best captured by an MEST. Each link $l_{i,j}$ is given a weight that is a function of normalized communication bandwidth, power, and latency:

$$\boldsymbol{\alpha} \times \boldsymbol{\psi_{i,j}} \times \left[\boldsymbol{\omega_{i,j}/max\ \_bw}\right] + (\mathbf{1} - \boldsymbol{\alpha}) \times \left[\boldsymbol{min\ \_latency/\lambda_{i,j}}\right] \quad (9)$$

where $\psi_{i,j}$, $\omega_{i,j}$, and $\lambda_{i,j}$ are link power consumption, link bandwidth, and link latency, respectively. We use separate values for the parameter α for FSOI links and electrical links due to their power consumption differences. The MRST structure for electrical links and MEST structure for FSOI links are constructed with the goal of minimizing the aggregate link weights, and an example of these structures is shown in Figure 65 (b)-(c). Note again that the routers are still not accurately mapped on the die during this stage, and we approximate virtual router locations at the center of each core. At the end of this step, all cores are connected with FSOI and electrical links.

7.3.4 CLUSTERING AND DUAL LEVEL ROUTER MAPPING

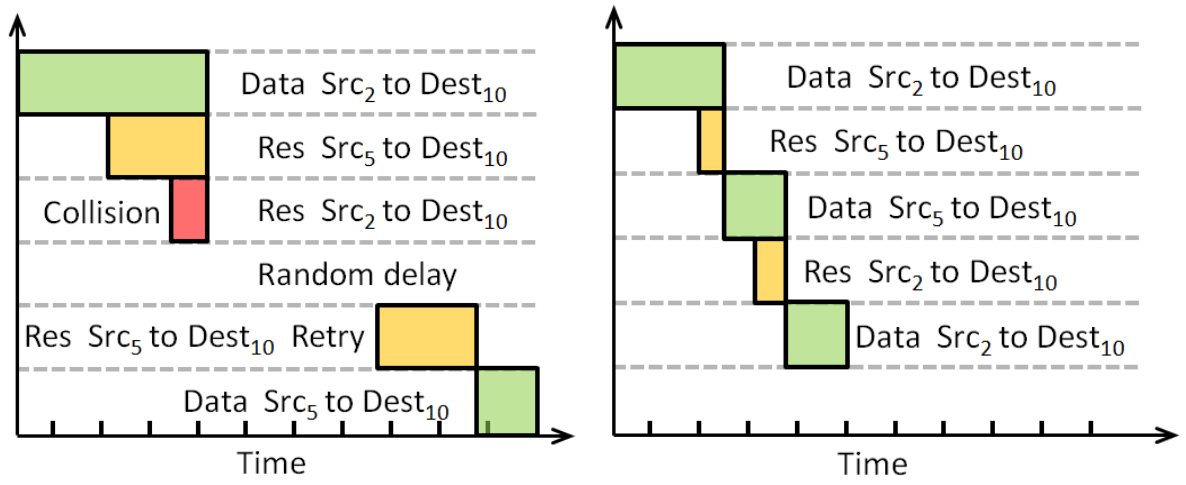
The objective of the subsequent clustering step is to merge the communication links in the MEST and MRST solutions; map hybrid nanophotonic-electric and electrical-only routers such that router counts are minimized and utilization of links and routers is improved; and optimize FSOI links.

We create a *heuristic* that computes connection strength between each node pair based on inter-node link bandwidth and power characteristics. Then starting with no edges between any nodes, we add edges in order of decreasing connection strength to create clusters, as shown in Figure 65 (d). The clusters are created utilizing a connection strength threshold such that intra-cluster short distance communication paths can be optimized utilizing electrical links and inter-cluster transmission can be performed using FSOI links. Each cluster represents a router in the final solution. But we still need to determine which communication flows will utilize FSOI links, electrical links, or a combination of both types of links. This problem is solved by using a *push-relabel maximum flow* algorithm. For every core n_i in the system, we create a corresponding pseudo core n_i^i, where inter-core communication for all *n* cores uses MEST links and all n'cores MRST links. The *n* and n'cores are linked with weights based on MRST and MEST links. Using the push-relabel maximum flow algorithm we generate a combined Steiner Tree and then merge the *n* and n'cores.

At the end of this stage, we utilize the *max-flow min cut algorithm* [47] to determine 1-hop or 2-hop routing for FSOI-based communication flows, to maximize bandwidth utilization while minimizing router resources. This process can also add or delete FSOI links as needed to meet any unsatisfied bandwidth or latency constraints; and tradeoff between performance and power requirements as discussed in next section.

7.3.5 PCR SIZE SYNTHESIS

In this step, we perform post processing of the combined MEST/MRST and router mapping to develop PCR regions. The root nodes in the MEST that include multiple and multi-depth branches are considered for integration into a PCR region with the nearest gateway interface router. More specifically, PCR regions cover nodes that are directly connected to the root nodes with *connection strength* lower than the links between the root nodes. For inter-PCR transfers, we set a size threshold M_{th} such that messages with size less than M_{th} transverse electrical links, while messages that exceed the threshold size travel to gateway interface routers and utilize FSOI links. Such a scheme ensures that small message size transfers do not encounter unnecessary E/O and O/E conversion delays, which would make their transfer over FSOI links less advantageous than over electrical links.



(a) **(b)**

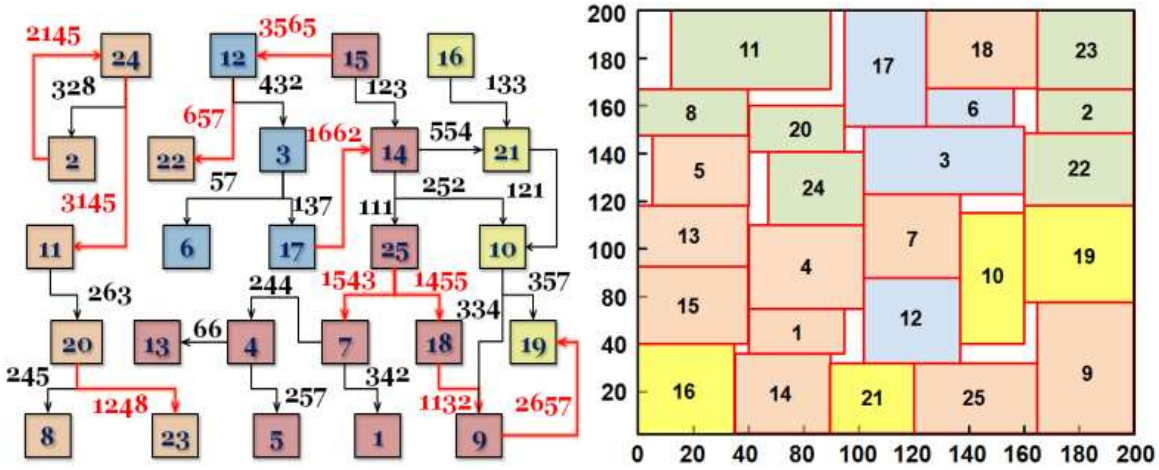
Figure 66 Scenarios for reservation channel collision (a) reservation process with FSOI collision (b) reservation process after adjusting serialization degree

7.3.6 CONFLICT ANALYSIS AND RESOLUTION

To reduce collision probability within the FSOI pipelined reservation channel, this final step attempts to minimize interference between various FSOI transactions. As we implement a pipelined router architecture with separate reservation channels, the reservation process can proceed while data transmission is in progress. Thus, more than one source core can attempt to reserve the same destination core, resulting in reservation collision (i.e., interference in modulated photonic links) at the destination node. This collision can produce erroneous data bits. Such collision can be detected using parity bits. In our architecture, transaction interference is avoided by managing link bandwidth via *modulation of the serialization degree*. Transactions from a source router (connected to the initiating core) to the sink router (connected to the target core) along each FSOI path are evaluated based on detailed communication schedules along a time-axis. In case of any conflicts between two transactions, we serialize these transactions such that both transactions can traverse the same router without interfering with each other.

Figure 66 summarizes this process. The channel reservation time is represented by the yellow colored horizontal bar. In the normal case when there is no collision, the reservation proceeds in parallel to data transmission. Once the reservation phase is complete, the next data transaction can begin. Figure 66(a) depicts a collision scenario with the red colored bar, where two reservation requests arrive in parallel with the first transaction's data transmission. This situation requires a reservation retry for the conflicting nodes after a specified retransmission delay, thus increasing latency. To eliminate this collision latency, our conflict analysis and resolution step utilizes serialization to modulate communication bandwidth such that multiple streams can coexist without collision. Figure 66(b) demonstrates how serialization can eliminate retransmission delays due to collision, thereby achieving overall lower transmission latency (at

the cost of a slight increase in area and power due to the need for serialization/deserialization circuitry).



(a) **(b)**

Figure 67 (a) Communication trace graph for multiple parallel applications, nanophotonic links in red color (b) custom layout with irregular topology

Table 11 MiBench Applications for Application Categories

Application category	*Applications*
Industrial	basicmath, bitcount, qsort, susan
Consumer	jpeg, lame, mad, tiff2bw, tiff2rgba
Office	ghostscript, rsynth, stringsearch
Networking	dijkstra, patricia
Security	blowfish, rijndael, sha

7.4 EXPERIMENTS

7.4.1 EXPERIMENTAL SETUP

We synthesized application-specific hybrid NoC architectures for multi-application workloads derived from five MiBench [173] benchmark categories: *(i)* Automotive and Industrial Control, *(ii)* Consumer, *(iii)* Office Automation *(iv)* Networking, and *(v)* Security. As the

MiBench benchmarks are written for a single processor, we created our own multithreaded implementation using Linux *pthreads*. To create multi-application workloads, we combined multiple MiBench applications executing in parallel, with execution priority assigned to each application in case of any contentions arising during accesses to memories or during task scheduling. We generated instruction and communication traces for the benchmarks via the Shade simulator [172]. Table 11 presents the 17 benchmarks across the five application categories that were considered. We implemented 5 multi-application workloads, corresponding to all applications available in each category, e.g., for the automotive and industrial control multi-application workload, we included parallel implementation of *(i)* basicmath, *(ii)* bitcount, *(iii)* qsort and *(iv)* susan benchmarks. In addition to MiBench benchmarks, we also evaluated our *HELIX* framework with *PARSEC* [137] application benchmark workloads. The Princeton Application Repository for Shared-Memory Computers (*PARSEC*) benchmark suite is composed of several multithreaded programs that represent next-generation shared-memory programs for SoCs.

Table 12 Communication Synthesis GA Parameter Ranges

Synthesis Parameters	***Range low***	***Range high***
Source Processor ID	1	$m \times n$
Destination Processor ID	1	$m \times n$
Generation Index	1	5
Number of Data Packets	1	1028
Electrical or Photonic Link	0	1
Number of hops (computed)	4	NA
Energy consumption (computed)	NA	As specified
Latency Constraints (specified)	0	As specified
Task ID	NA	NA

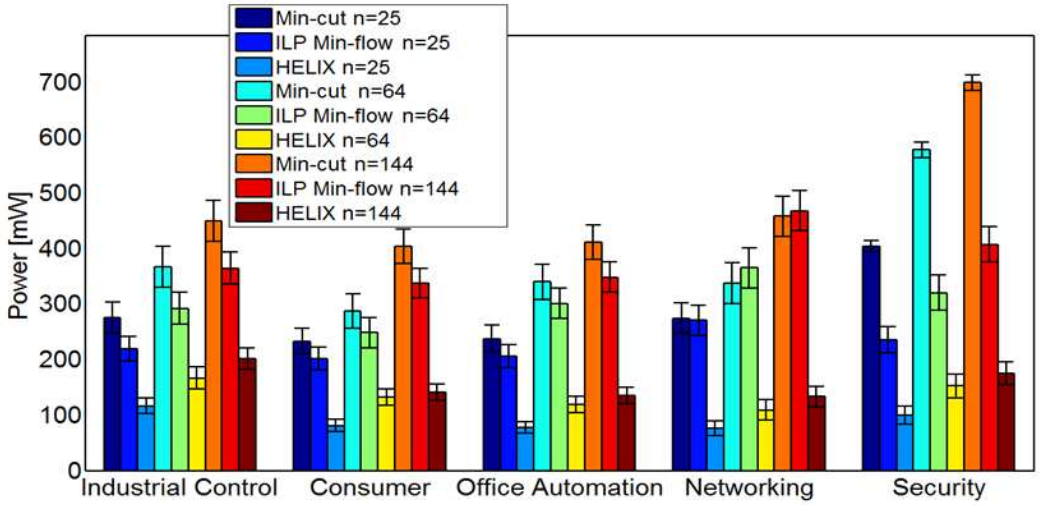
Our core mapping GA *chromosome* consists of parameters with ranges as defined in Table 12. For our core mapping GA initial population size included 2000 randomly generated application mappings. We evaluated a range of crossover mutation and probabilities and ultimately utilized probability values of *0.34* and *0.42* respectively. The best fitness value *chromosome* in each iteration which resulted in minimum power consumption while meeting performance constraints was cached to prevent being overwritten by a non-dominated solution chromosome. As the GA is a stochastic search algorithm, it is not possible to formally specify convergence criteria based on optimality, therefore we terminated our GA when the best solution quality did not change over a predefined number of iterations (2000).

Figure 67(a) shows the enhanced communication trace graph (CTG) of 4 industrial applications running in parallel, which is generated after running the GA algorithm. Vertices in the CTG represent cores on which tasks have been mapped. Note the initial link selection that allocates some flows to electrical links and others to FSOI links. Communication flows with stringent bandwidth and/or latency demands generally get mapped to the more efficient FSOI links in this first step. But note that this initial assignment can be modified as the solution is refined in the later steps of the *HELIX* design flow. Figure 67(b) shows the floorplanning solution for this CTG graph, where cores of each application are depicted by a separate color.

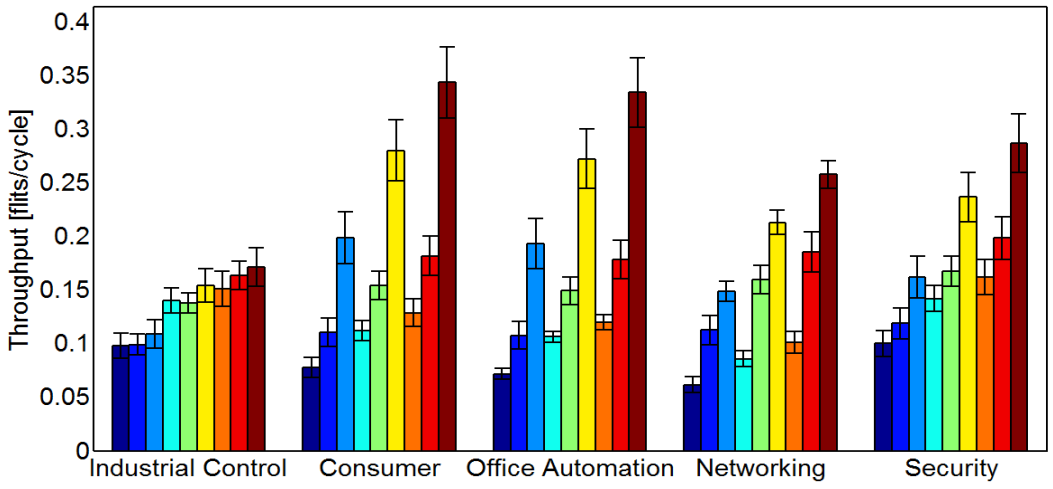
During floorplanning, we set weighed communication power constant $\propto$ and link distance constant β values at 0.5 each based on experimental analysis. We utilized a public domain GeoSteiner3.1 Steiner Tree solver [174] to generate the MEST and MRST networks and utilized the *lp_solve* optimizer [175] to solve the Mixed Integer Linear Programming (MILP). We set weight values for $\propto$ to *0.46* and *0.68* respectively, during MEST generation for electrical links and MRST generation for FSOI links. We created clusters utilizing a *0.38* connection strength

threshold to optimize electrical intra-cluster short distance communication and inter-cluster transmission using FSOI links. Based on experimental analysis, we set a normalized M_{th} threshold of *0.33* in PCR regions such that communication messages less than the size of M_{th} transverse through electrical links and the messages that exceed the size of the threshold travel through FSOI links. We combined the various components of our *HELIX* framework using a python scripting interface [176].

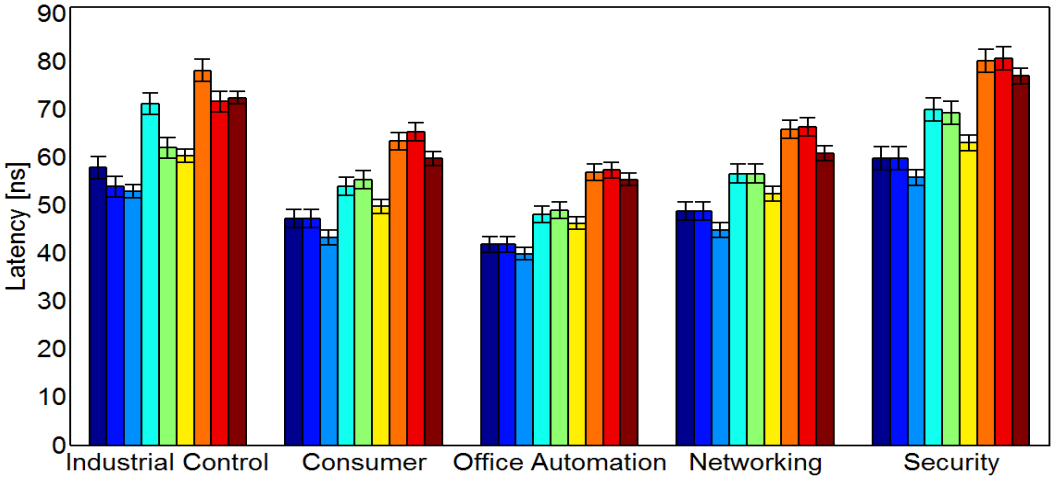
The static and dynamic power consumption of electrical routers as well as the power consumption for optimally sized repeated Cu wires is obtained from a modified version of the Orion 2.0 simulator [141]. Our synthesis process targeted the *32 nm* node technology and utilized a *400* mm^2 SoC die area. The delay of an optimally repeated and sized electrical (Cu) wire at 32 nm was assumed to be *42 ps/mm* [36]. Photonic free space communication delay is *3.3456 ps/mm* (compared to photonic waveguide delay of *15.4 ps/mm* [36]) requiring much less buffering during transfers compared to traditional electrical or waveguide based hybrid photonic NoCs. The intrinsic speed of a MQW is practically limited by the driver electronics and the well-known quantum-confined Stark effect working at sub-picosecond time scales. We modeled a $1\mu m$ thick *5V* modulator with $10\times10\mu m^2$ area and capacitance of *11 fF*, calculating the per cycle electrical energy of the device as *140fJ* which is in line with prior estimates [177]. Our implementation assumed modulator driver delay of *9.5 ps*, modulator delay of *3.1 ps*, photo detector delay of 0.*22 ps* and receiver delay of *4.9 ps* [44]. Our hybrid NoC with an irregular topology was modeled at the cycle accurate granularity by extensively modifying our in-house cycle accurate SystemC-based [158] NoC simulator derived from the Noxim [133] simulator.



(a)



(b)



(c)

Figure 68 Synthesis result comparison (a) power (b) throughput (c) latency

As no prior published work exists on synthesizing custom application-specific hybrid nanophotonic-electric NoCs for comparison, we compared our synthesis results with respect to application-specific NoC synthesis frameworks in [45] and [47] that synthesize purely electrical NoCs. We implemented the floorplan-aware design process in [47] that accounts for wiring complexity and detects timing violations on the NoC links early in the design cycle. This algorithm was implemented in two phases. Within the first phase, we selected a topology that best optimizes user objectives satisfying all design constraints, and in the second phase we varied a number of design parameters such as NoC clock frequency and link width to find a solution that best optimized all design constraints. Similarly, we implemented the two-stage synthesis methodology as presented in [47] which consist of core to router mapping and custom topology and route generation.

We evaluated various SoC complexities during our experimental analysis to better understand the impact and scalability of our *HELIX* synthesis framework for small (25 cores), medium (64 cores) and large (144 cores) sized SoCs, when compared to the frameworks in [45] and [47].

7.4.2 EXPERIMENTAL RESULTS

This section analyzes the hybrid nanophotonic-electric NoC designs synthesized by our *HELIX* framework for the various multi-applications workloads. The results of synthesis for the 25, 64, and 144 node complexity SoC platforms are shown in Figure 9, for the five multi-application MiBench workloads. Our *HELIX* synthesis framework provides on average 2.82×, 3.12× and 3.49× reduction in power for the 25, 64, and 144 core SoC platforms respectively compared to application specific electrical NoC utilizing approaches from [45] and [47]. This

improvement in power dissipation with *HELIX* relative to [45] and [47] is a result of: *(i)* congestion reduction in the electrical links due to offloading of a large portion of the global communication to FSOI links; *(ii)* reduction in electrical link switching activity; *(iii)* shorter link lengths; and *(vi)* smaller buffer resources compared to electrical-only application-specific NoC architectures synthesized by [45] and [47]. Due to the use of fast and high bandwidth FSOI links as well as reduced congestion in the electrical NoC, the communication latency improved with *HELIX by 1.18×, 1.23× and 1.25×* and throughput by *1.68×, 1.69× and 1.78×* for the 25, 64 and 144 core SoC platforms.

We observed that *HELIX* was able to achieve a significant reduction in the number of gateway interface routers through clustering and dual level router mapping, with as few as 33% gateway interface routers compared to the router count before clustering. The dual level router mapping step also adds paths enabling inter-cluster long distance global communication using 2-hop FSOI links to minimize power. These additions allow the electrical NoC router count and complexity to be reduced compared to results obtained from [45] and [47]. We also observed that the conflict analysis and resolution step in *HELIX* reduced MQW modulator and detector counts by approximately 50% by intelligent management of serialization degrees at the nanophotonic-electric interfaces.

Our *HELIX* synthesis framework was able to achieve a viable solution for all application workloads and SoC complexities that we evaluated. Each stage in our synthesis framework worked seamlessly, complementing each other to balance conflicting requirements to solve the nontrivial problem of synthesizing application-specific hybrid free-space photonic-electric NoC fabrics. Table 13 summarizes key synthesis parameters for the 25, 64 and 144 SoC sizes, for the MiBench multi-application workloads. It is interesting to observe that the number of photonic

free-space gateway interfaces is significantly lower than the number of electrical routers, and accounts for only 20% of all routers. This is in contrast with previously proposed [36] [37] [38] [39] [40] [41] nanophotonic architectures. The number of clusters and gateway interfaces correlates well with each other and by judiciously selecting FSOI hop counts, the framework minimizes area and the number of modulators and photodetectors required, without violating performance constraints.

HELIX is also able to reduce the average number of hops in the electrical network by up to 4×, electrical link area by up to 1.24×, and link lengths for the electrical network by up to 2.67× compared to previously published electrical NoC synthesis techniques in [45] and [47]. This improvement is possible due to the *HELIX* floorplanner placing cores communicating via FSOI links farther apart and the cores communicating via electrical links closer, achieving two fold benefits by replacing long distance electrical links with more efficient FSOI links and placing cores closer that communicate with electrical links.

The results of synthesis for the 25, 64, and 144 node complexity SoC platforms are shown in Figure 71, for twelve multi-threaded *PARSEC* benchmarks (*blackscholes (bl), bodytrack (bo), canneal (ca), dedup (de), facesim (fa), ferret (fe), fluidanimate (fl), freqmine (fr), streamcluster (st), swaptions (sw), vips (vi), x264 (x2)*). Once again, for the *25, 64* and *144* core SoCs, *HELIX* achieves a notable power dissipation improvement of *3.28×, 3.40×, 2.58×* respectively, compared to the results obtained from the synthesis frameworks in [45] and [47], as well as improvements in throughput by *1.11×, 1.14×, and 1.16×* and average transfer latency *by 1.44×, 1.56×, and 1.49×* respectively.

The breakdown of normalized power consumption in Figure 70 for the *PARSEC* benchmarks demonstrates how *HELIX* can effectively improve power consumption in all

categories by managing nontrivial trade-offs during the synthesis process, balancing transfers across electrical and photonic planes to provide superior results. The lower buffer power can be attributed to lower latency in free space paths, allowing for routers with less buffer space within the electrical network. Link power consumption accounts for majority of improvements due to the free space photonic path utilization consuming lower power that also reflects in lower electrical network area overhead as shown in Figure 69.

Table 13 Comparison of Synthesis Parameters

P =Average Power improvement compared to [45] and [47], PR = Number of photonic routers, C = Number of clusters, PR=Max PCR Size, EH=Max Electrical Hop Count, PH=Max Photonic Hop Count, SD=Serialization Degree

Application	*P [45]*	*P [47]*	*PR*	*C*	*PR*	*EH*	*PH*	*SD*
25 core NoC								
Industrial	2.363	1.883	7	7	4	4	2	2
Consumer	2.880	2.491	6	6	3	3	2	3
Office	3.061	2.663	8	8	3	4	2	3
Networking	3.603	3.551	8	6	4	3	1	4
Security	4.062	2.367	7	6	5	4	1	4
64 core NoC								
Industrial	2.198	1.751	16	15	4	4	2	3
Consumer	2.175	1.879	17	17	5	4	2	3
Office	2.867	2.532	16	16	5	4	2	4
Networking	3.098	3.352	18	16	4	4	2	3
Security	3.796	2.104	18	16	4	4	2	4
144 core NoC								
Industrial	2.226	1.806	28	25	6	4	2	3
Consumer	2.847	2.377	25	23	7	5	2	4
Office	3.044	2.577	26	24	6	5	2	3
Networking	3.437	3.519	25	28	6	4	2	4
Security	3.993	2.327	26	25	7	4	2	4

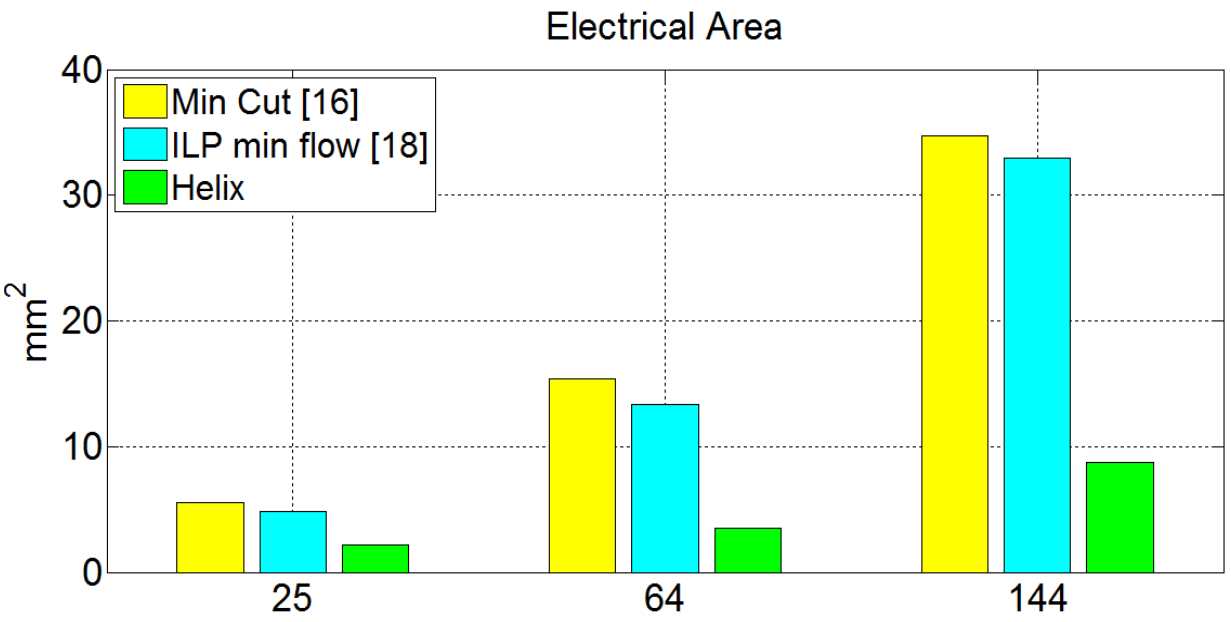


Figure 69 Area overhead comparison for *HELIX* synthesized electrical network for *PARSEC* application benchmarks

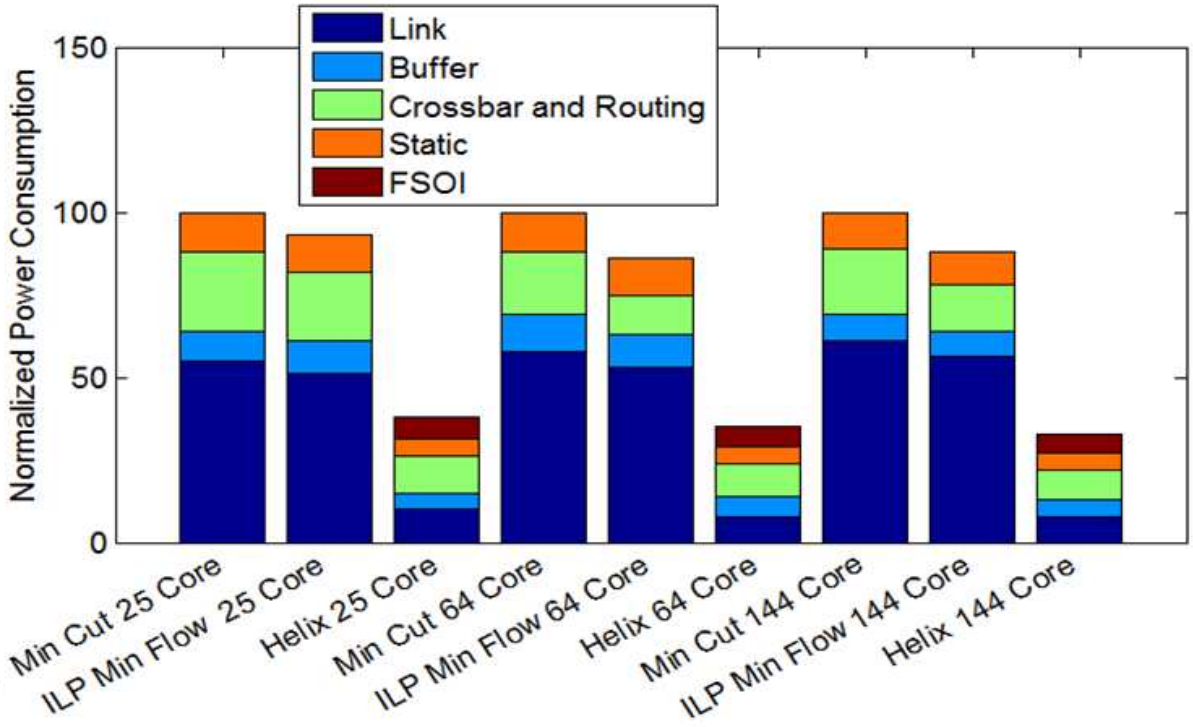
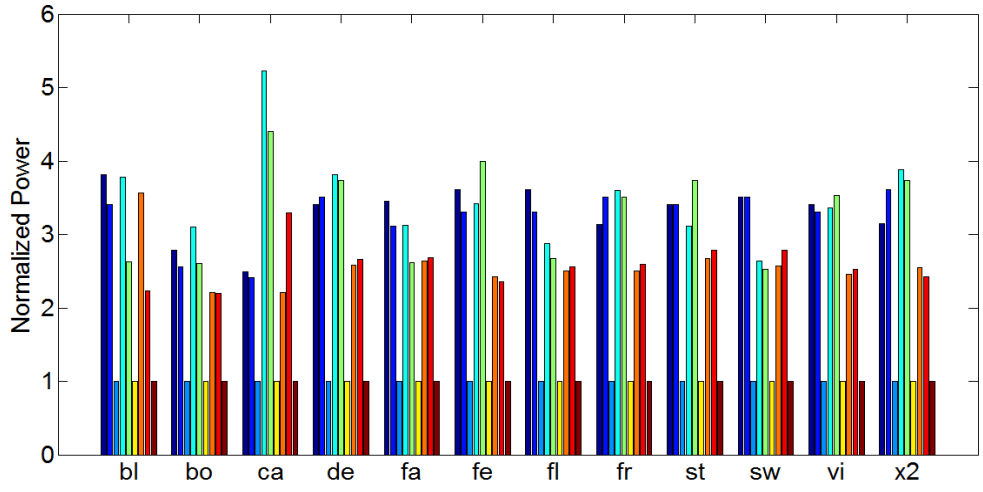
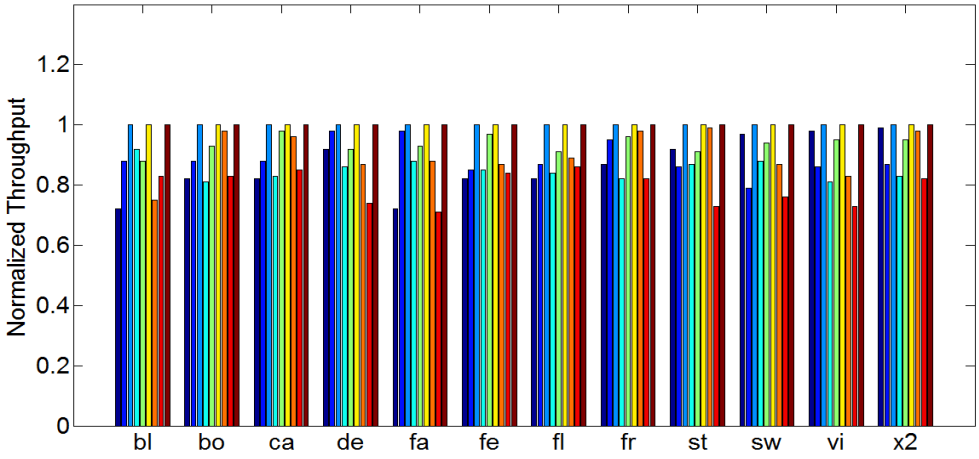


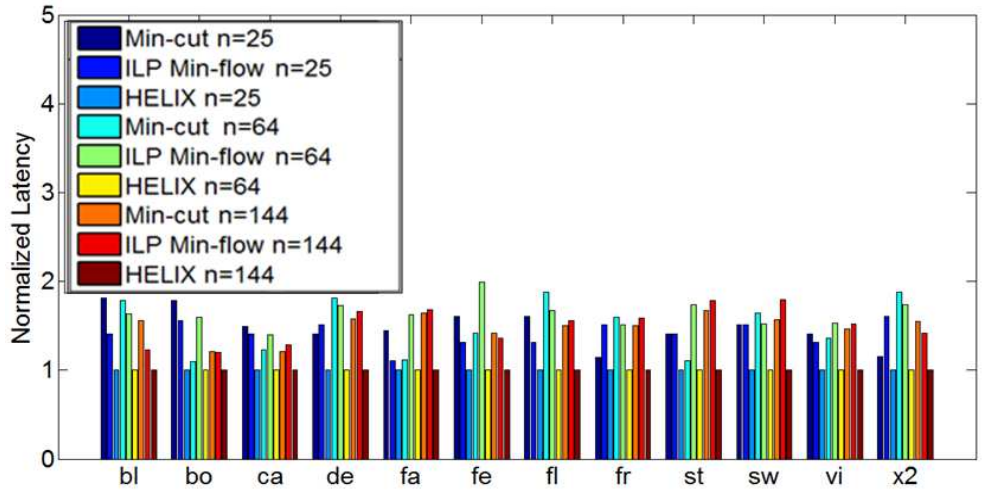
Figure 70 Normalized breakdown of power consumption



(a)



(b)



(c)

Figure 71 Synthesis result comparison for *PARSEC* multi-threaded workloads (a) average power (b) throughput (c) average latency

7.5 RESULT SUMMARY

In this chapter, we presented the *HELIX* framework to *synthesize application specific hybrid nanophotonic-electric NoCs with irregular topologies*. To the best of our knowledge this problem has not been address before in any prior work. Based on our experimental studies, we demonstrate that the proposed techniques in the *HELIX* framework produce a superior NoC architecture that satisfies all performance requirements for *MiBench* multi-application workloads and *PARSEC* multi-threaded workloads, while achieving an average of *3.06*× reduction in power dissipation across SoC platforms of varying complexity, compared to previously proposed application-specific electrical-only NoC synthesis frameworks. By addressing the many challenges related to the overheads of microring resonators and photonic waveguide based architectures, we also propose a practical framework aimed at bringing hybrid nanophotonic-electric NoCs based on FSOI links and electrical links closer to reality.

8 3D-HELIX: DESIGN AND SYNTHESIS OF HYBRID FREE SPACE APPLICATION-SPECIFIC 3D NOC ARCHITECTURES

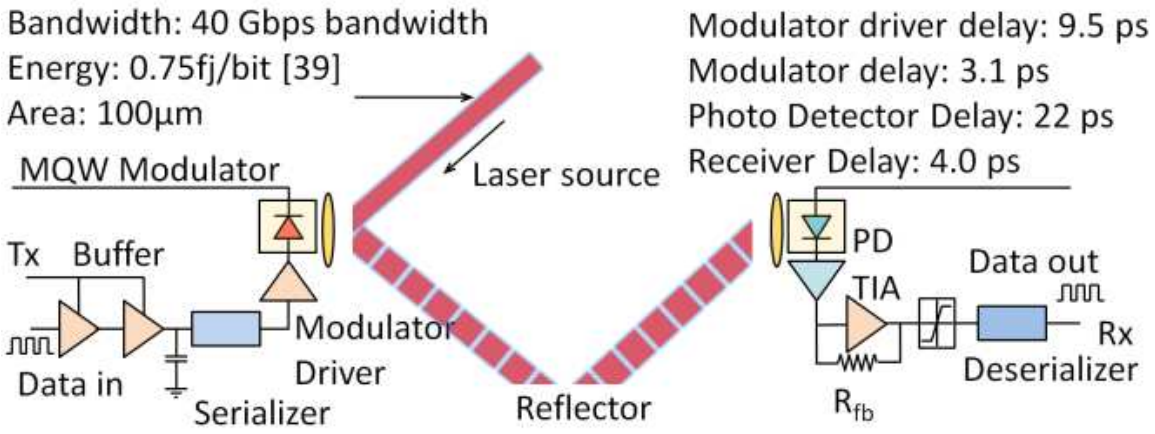
With the advent of 3D chip stacking technology, application-specific and heterogeneous three dimensional chip multi-processors (3D CMPs) are projected to become the building blocks of future parallel processing systems. In such 3D CMPs, network-on-chip (NoC) architectures will enable communication between multiple heterogeneous cores. However, NoCs face several challenges, including limited bandwidth, high latency, and high power dissipation. Hybrid nanophotonic-electric NoCs are being considered as a solution to address the above challenges due to their desirable performance and power characteristics. These emerging communication architectures require substantial optimization to realize their full potential. Optimizing hybrid nanophotonic-electric 3D NoCs requires intelligent traversal through a massive design space, which is non-trivial. *No prior work has addressed the problem of synthesizing and optimizing application-specific hybrid nanophotonic-electric 3D NoCs with an irregular topology to connect heterogeneous cores on a 3D CMP*. Considering the above unaddressed major challenge, in this chapter we propose a synthesis framework called *3D-HELIX* that can optimize application-specific hybrid nanophotonic-electric 3D NoCs.

8.1 MOTIVATION FOR 3D INTEGRATION

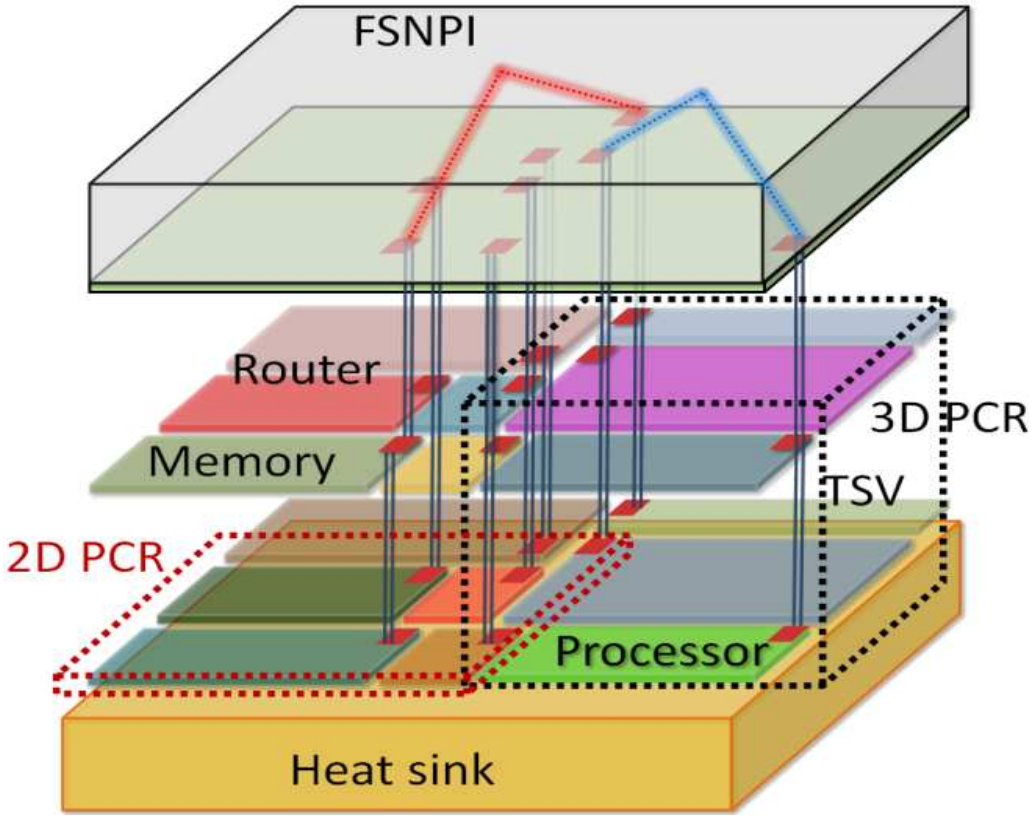
With the push towards integrating more and more cores on a die to enhance parallel processing capabilities, 3D integration has emerged as an interesting way to achieve high core densities on a chip. Wafer-to-wafer bonded 3D integrated circuits (3D-ICs) place active devices (processors, memories) within multiple active layers and vertical Through Silicon Vias (TSVs) connect cores across the stacked layers. Multiple active layers in 3D-ICs can enable increased

integration of cores within the same area footprint as traditional single layer 2D-ICs. In addition, long global interconnects between cores can be replaced by shorter inter-layer TSVs, improving performance and reducing on-chip power dissipation. Recent 3D-IC test chips from IBM [32], Tezzaron [33] and Intel [178] have confirmed the benefits of 3D-IC technology.

Electrical network-on-chip (NoC) communication fabrics are commonly used in 2D processing chip architectures to connect various heterogeneous cores together. These fabrics are however severely constrained due to their long multi-hop latencies, low bandwidth density, and high power dissipation [89]. In 3D-ICs that utilize 3D NoC fabrics, the fundamental power, delay, and noise susceptibility limitations of traditional copper (Cu) interconnects are still severe. To overcome these limitations, alternative interconnect materials are needed. Photonic interconnects [29] represent one promising emerging solution that can replace Cu interconnects on a chip and help overcome their latency, bandwidth, and power bottlenecks. Photonic interconnects can transfer data with much more energy efficiency than Cu interconnects especially over long distances across a chip. Thus on-chip photonic interconnects are being



(a)



(b)

Figure 72 Building blocks of free-space on-chip photonic interconnects: (a) modulator and receiver circuit (b) 3D integration of electrical and FSNPI layer interconnect including photonic concentration region (PCR)

actively explored as a promising alternative to Cu interconnects for global communication, allowing data to be transferred across a chip at a much faster light speed and with power dissipation that is independent of link length [89].

8.2 BACKGROUND: FSNPI ARCHITECTURE

To overcome challenges with waveguides and silicon microring resonators, on-chip free-space nanophotonic interconnects (FSNPIs) have recently been proposed [42] [43]. Dense Multiple Quantum Well (MQW) devices are used for electro-optic modulation, consuming less than *1* pJ/bit energy. These MQW devices can be configured either as absorption modulators or photo-detectors (PDs). On-chip optical interconnects utilizing MQWs can operate at 40 Gbps

bandwidth [31] to instantiate single-hop point-to-point or multi-hop transfers through free-space optical links. Most interestingly, MQW modulators do not suffer from thermal tuning challenges of silicon microring resonators 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 and are better suited for high-density optical interconnects due to their small active area and improved misalignment tolerance. 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.

Figure 72*(a)* summarizes the building blocks of FSNPIs with MQW modulators and PDs, and its system level integration with an electrical NoC fabric. Serializer/deserializer circuits enable trade-offs between communication power, area and bandwidth by reducing photonic components through higher serialization degree. A unique feature of our hybrid nanophotonic-electric NoC fabric shown in Figure 72*(b)* is the reconfigurable traffic partitioning between electrical and photonic links. To minimize implementation costs, our synthesis framework limits the number of gateway interfaces between the electrical and photonic layers. An adaptive photonic concentration region (PCR) ensures appropriate scaling and utilization with changing communication demands. A PCR is defined as the number of cores around the gateway interface that can utilize the FSNPI path for communication. Cores within the same PCR communicate with each other via the electrical NoC (intra-PCR transfers). Cores that need to communicate and reside in different PCRs communicate using photonic paths (inter-PCR transfers). The electrical NoC transfers use XYZ routing, and a modified PCR-aware routing scheme for selective data

transmission through the FSNPI links, with timeout-based regressive deadlock handling, based on the approach presented in [36].

As shown in Figure 72 (b), we consider a heterogeneous CMP platform with a dedicated photonic layer that supports FSNPI links, interfacing with an electrical NoC. Our electrical NoC is composed of two types of routers: *(i)* conventional four stage pipelined electrical routers that have *n* I/O ports and interface with local cores; and *(ii)* hybrid gateway interface routers (Figure 73) that are also four-stage pipelined but have additional photonic ports (a total of *n+3* I/O ports). The photonic link interface in gateway routers is responsible for sending/receiving flits to/from photonic links in the photonic layer. Both types of routers have an input and output queued crossbar with a 4-flit buffer on each input/output port, with the exception of the photonic ports in gateway interface routers that use double buffering to cope more effectively with the higher photonic path throughput. The serializer/deserializer modules can support serialization degree of 2, 4, 8 and 16, and allow for very high optical I/O pad density. The resulting high-density 2D array of surface-normal optoelectronic MQW devices can provide the necessary intra chip bandwidth density without the complexity of wavelength division multiplexing (WDM) [171].

Our arbitration approach is different from FSNPI-based gateway interface routers proposed in [42] and [43] that utilize transfers without any arbitration. These routing schemes directly stream data to destination cores and manage collision of photonic data with a collision handling scheme (e.g., when multiple source nodes send data to the same destination core). But we observed that the performance benefits of eliminating arbitration are overshadowed by high penalties of collision handling and retransmission for high performance communication flows. Therefore in our gateway interface routers we implemented support for reservation channels to

reserve FSNPI data paths. An additional input and output reservation channel port is added to the routers for this purpose.

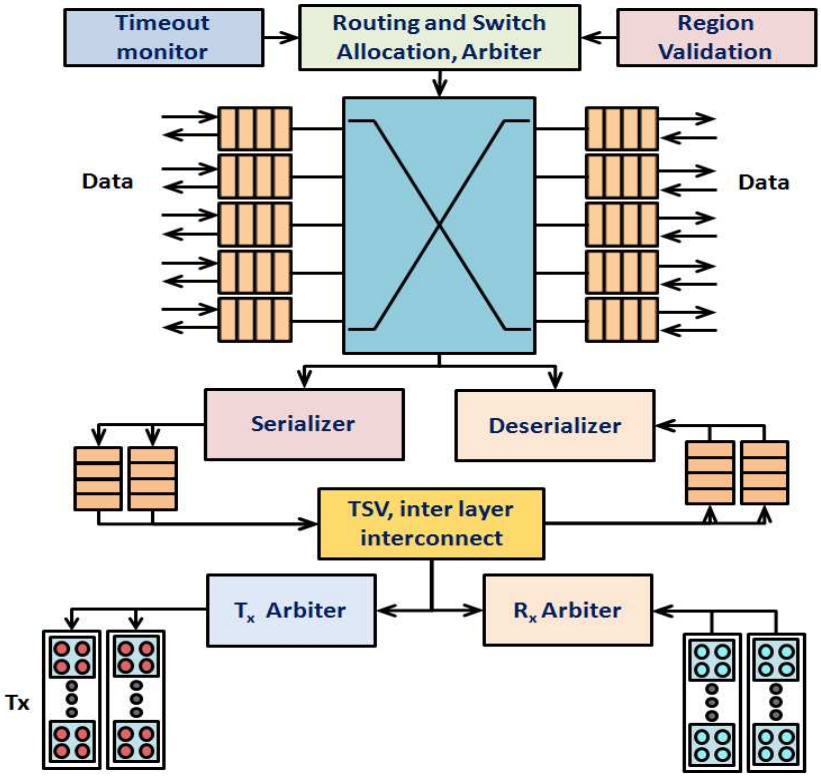


Figure 73 3D Gateway interface FSNPI router architecture

Finally, our FSNPI architecture supports multiple hops to balance performance and power goals. For a 1-hop $(m \times n \times l)$ path with flit width of k, each node on the 3D CMP needs $2\times k[(m \times n \times l) - 1]/(MWQ$ Gbps/(GHz CPU clock)) MQW devices, while a 2-hop $(m \times n \times l)$ path with the same flit width needs $4\times k[(m + n + l) - 2]/(MWQ$ Gbps/(CPU Clock)) MQW devices [43]. As an example, for a 3D CMP with 128 cores and 2 active layers, a 1-hop FSNPI based hybrid 3D NoC with flit width of 256 bits at 40 Gbps/link and a 3.88 GHz CPU clock requires *6308* MQW devices. These photonic components for a *20mm×20mm* CMP die size will consume < *4.2* mm^2 on-chip area for a 1-hop FSNPI-based NoC with *100*μm MQW devices. In contrast, a 2-hop FSNPI based hybrid 3D NoC will require only *1590* MQW devices with < 1

mm^2 area and a 5.0× power reduction over a 1-hop NoC, but at the cost of system bandwidth drop from *300 to 45* Tbps. We explore hop-count selection on a per-communication flow basis to enable power-bandwidth trade-offs in our 3D-HELIX synthesis framework.

8.3 PROBLEM FORMULATION

This section presents an overview of our problem formulation for synthesizing application-specific heterogeneous 3D hybrid nanophotonic-electric NoC architectures with 3D-HELIX:

8.3.1 APPLICATION WORKLOAD CONSTRAINTS

- Application communication trace graph *G(V,M,L)* for each application in a multi-application workload, where $v_i \in V$ is a set of processing cores, $m_i \in M$ a set of memory blocks, $l_i \in L$ a set of directed communication links;
- Application-specific communication bandwidth constraints $\omega_{i,j}$ in bits/cycle and latency constraints $\lambda_{i,j}$ in cycles between $\{v_i, v_j\}$ or $\{m_i, m_j\}$;

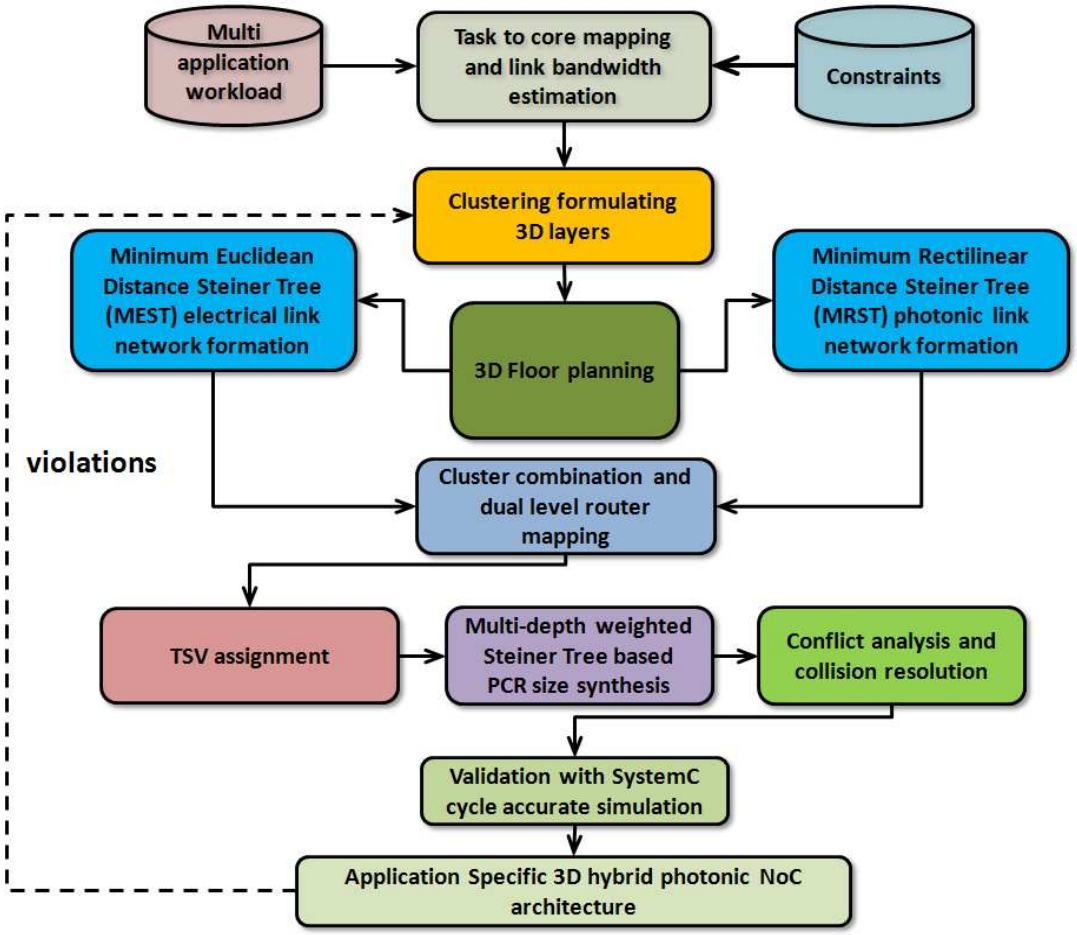


Figure 74 *3D-HELIX* hybrid nanophotonic-electric NoC synthesis flow

8.3.2 SOC PLATFORM CONSTRAINTS

- X_{max}, Y_{max} and Z_{max} are the maximum dimensions of the die along the *X*, *Y* and *Z* axes; and the aspect ratio X_{die}/Y_{die} of the synthesized die should be between *0.9-1.1* to obtain an approximately square die layout;
- Each network link is constrained by a maximum length γ that represents the maximum distance a signal can travel in a single cycle, based on CMOS process technology;
- Single layer FSNPI network constrained by on-chip area and communication requirements;

8.3.3 PROBLEM OBJECTIVE

- Synthesize a 3D hybrid nanophotonic-electric application-specific heterogeneous NoC architecture $J(R, L_e, L_p, C)$ where R is a set of hybrid (photonic-electric gateway interface) routers, L_e and L_p represents the set of electrical and photonic links, and C is a core-to-die mapping function; such that communication power is minimized while meeting bandwidth and latency constraints of the given application(s), and platform constraints of the SoC.

8.3.4 CONFIGURATION PARAMETERS

- Application task to core mapping;
- Mapping and layout of cores and memories for each active layer;
- Number and layout of hybrid electro-photonic and electrical-only routers that utilize a set of photonic $p_i \in P$ and/or electrical links $e_k \in E$ to support communication for a given multi-application workload;
- Sizes of photonic concentration regions (PCRs) that determine the cores/memories allowed to use each hybrid photonic routers on the die (see Section 3 for details);
- Serialization degree D_n at electro-photonic interfaces;
- Hop count (1-hop or 2-hop) selection for FSNPI links;

8.4 *3D-HELIX* SYNTHESIS FRAMEWORK OVERVIEW

In this section, we present our framework for synthesizing hybrid nanophotonic-electric 3D NoCs for heterogeneous CMPs. The framework consists of the following steps as shown in Figure 74 *(i)* task-to-core mapping; *(ii)* formulating 3D layers; *(iii)* floorplanning; *(vi)* Steiner

tree based network formation; *(v)* link clustering and dual level router mapping; *(vi)* TSV assignment*; (vii)* PCR allocation; *(vii)* conflict analysis and resolution; and *(viii)* validation with cycle-accurate simulation. Due to lack of space, here we briefly discuss each step. The following subsections provide an overview of these steps.

8.4.1 TASK TO CORE MAPPING

In this first step, we perform task to core mapping, scheduling, and link bandwidth estimation. The step involves mapping of *n* tasks to *m* heterogeneous cores for the given application(s) task flow graph. We perform task execution-time estimation as well as estimation of inter-core data transfers using an instruction simulator [172]. We implement a genetic algorithm (GA) [170] to accomplish task to core mapping. The GA chromosome consists of possible mappings for the given application tasks to available cores, as well as virtual link type (electrical or photonic) between cores to satisfy bandwidth and latency constraints at a coarse granularity. The GA cost function represents overall communication power and the GA attempts to create a mapping and task schedules to minimize this power.

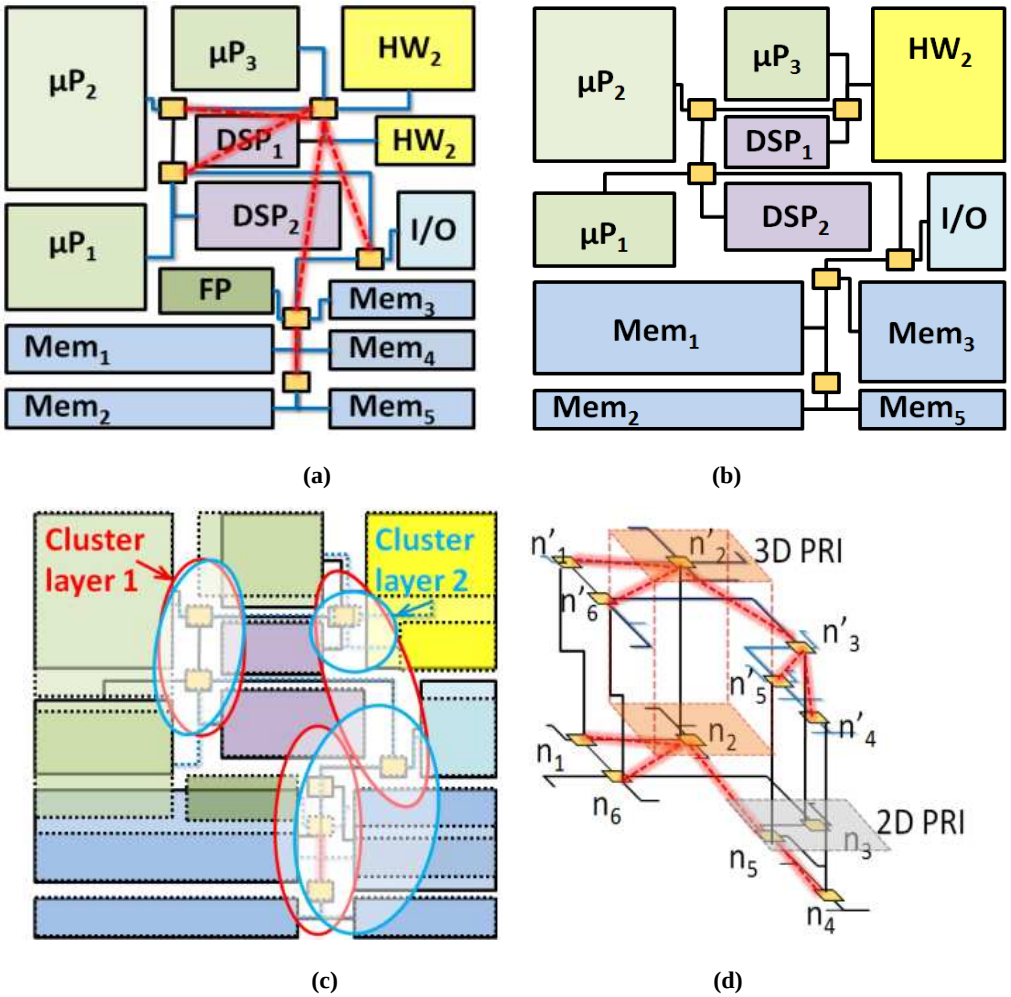


Figure 75 (a) Layer one, Steiner tree (MEST) for electrical network and minimum rectilinear distance Steiner tree (MRST) for FSNPI links (b) layer two, MEST and MRST (c) clustering for dual level router mapping (d) TSV assignment and PCR generation

8.4.2 CLUSTER FORMULATION FOR 3D LAYERS

As a planar 2D electrical NoC is efficient for transmission of small length messages, whereas interlayer and FSNPI links can provide significant benefits for large packet length messages, we implement a *k-means* clustering algorithm to partition cores to dies based on their communication characteristics. Here k-stands for number of clusters that represents number of layers within the 3D CMP. We utilized sum of weighed bandwidth squares as a minimization function. The major constraint was to ensure that the sum of areas of cores for each die was less

than the planar die area. This constraint allows a 2D floorplanner to work seamlessly to position cores on each die. The sum of weighted bandwidth squares within a cluster is minimized so that horizontal communication bandwidth in each layer is minimized while the vertical communication bandwidth is maximized.

8.4.3 FLOORPLANNING

In the next step, floorplanning is performed within each cluster obtained by *k-means* clustering for each 3D layer, to determine more precise placements for cores in dies. This step also influences the network topology and therefore must be cognizant of the communication between cores during floorplanning. In general, communication delay and power consumption play a significant role in determining the optimal network topology for an application. With hybrid NoC architectures, the floorplanning step becomes more complicated as the power consumption and delay of electrical wires and FSNPI links differ significantly. As per our best knowledge there is no floorplanning tool available that can support such hybrid FSNPI and electrical wire based network architectures during floorplanning. We therefore designed an enhanced system-level NoC floorplanning tool that uses mixed integer linear programming (MILP) to perform core placement on a die. Our MILP minimization objective function is a linear combination of the weighed communication power-latency and overall chip area, representing the metrics optimized in this step:

$$\left[\sum_{\forall\, c(u,v)\in E}^{i,j} l(u,v) \times Cp_{i,j}\right] \propto +[X_{max} + Y_{max}]\beta \qquad (10)$$

where, $l(u,v) \times Cp_{i,j}$ is the weighed communication power and link distance between cores, $\propto$ and β are constants, and X_{max} and Y_{max} are the dimensions of the die along the X and Y axes. During the floor planning process we also define unity aspect ratio X_{max}/Y_{max} as a constraint to obtain an approximately square shaped floorplan. As FSNPI links consume less power than electrical links, the floorplanner allows placing cores communicating via FSNPI links farther apart than cores communicating via electrical links. The floorplanner also works independently for each layer. Note also that at this stage, routers have yet to be allocated, and are assigned arbitrary locations with respect to cores (1 virtual router / core). The output of this step is a floorplan as shown in Figure 75(a) and (b).

8.4.4 MEST AND MRST BASED NETWORK FORMATION

In this step, we initiate the network formation by generating Rectilinear and Euclidian Minimum distance Steiner Trees, where the link weight for link $l_{i,j}$ is a function of normalized communication bandwidth, power, and latency:

$$\boldsymbol{\alpha \times \psi_{i,j} \times \left[bw_{i,j}/max\ \_bw\right] + (1-\alpha) \times \left[min\ \_latency/latency_{i,j}\right]} \qquad (11)$$

where $\psi_{i,j}$, $bw_{i,j}$, and $latency_{i,j}$ are link power consumption, link bandwidth, and link latency, respectively. We use separate weight values (α) for FSNPI links and electrical links due to their power consumption differences. We ultimately generate a Minimum Euclidean Distance Steiner Tree (MEST) for electrical links and a Minimum Rectilinear Distance Steiner Tree (MRST) for FSNPI links, as shown in Figure 75 (a)-(b). Note again that the routers are still not accurately

mapped on the die during this stage, and we approximate virtual router locations at the center of each core. Finally, we connect all cores (virtual routers) utilizing FSNPI and electrical links.

8.4.5 CLUSTERING AND DUAL LEVEL ROUTER MAPPING

The objective of the subsequent clustering step is to merge the communication links in the MEST and MRST solutions, and map conventional and hybrid routers such that router counts are minimized and utilization of links and routers is improved. This step considers tradeoffs between local and global communication assignment to electrical or FSNPI links. We developed a heuristic that computes *connection strength* between each node pair on the same layer based on link bandwidth and power characteristics. Then starting with no edges between any nodes, we add edges in order of decreasing connection strength to create clusters, as shown in Figure 75 (c). We repeat this process for each layer. The clusters are created utilizing a connection strength threshold such that intra-cluster short distance communication paths can be performed by utilizing electrical links and inter-cluster transmission can be performed using FSNPI links.

Each cluster represents a router in the final solution. But we still need to determine which communication flows will utilize FSNPI links, electrical links, or a combination of both. This is accomplished using a *push-relabel maximum flow* algorithm. For every core n_i in the system, we create a corresponding pseudo core n_i^i, where inter-core communication for all *n* cores uses MEST links and all n'cores uses MRST links. The *n* and n'cores are linked with weights based on MRST and MEST links as shown in Figure 75 (d). Using the *push-relabel maximum flow algorithm* we generate a combined Steiner Tree and then merge the *n* and n'cores. At the end of this stage, we utilize *the max-flow min cut* algorithm to determine 1-hop or 2-hop routing for

FSNPI-based communication flows, to maximize bandwidth utilization while minimizing router resources.

8.4.6 TSV ASSIGNMENT

In this step, we assign vertical TSV electrical links and FSNPI links for inter-layer communication. This is achieved by again applying the *push-relabel maximum flow* algorithm. For every core n_i in the system, we create a corresponding pseudo core n_i^i, where inter-layer communication for all *n* cores uses electrical links and all n'cores uses FSNPI links. Using *the push-relabel maximum flow* algorithm we generate a combined interconnect in 3D layers and then merge the *n* and n'cores. Similarly to the intra-layer mapping, we utilize *the max-flow min cut* algorithm to determine 1-hop or 2-hop routing for FSNPI-based inter-layer communication flows, to maximize bandwidth utilization while minimizing router resources. This process also can add or delete FSNPI links as needed to meet any unsatisfied bandwidth or latency constraints and balance performance and power requirements based on 1-vs-2 hop routing trade-offs.

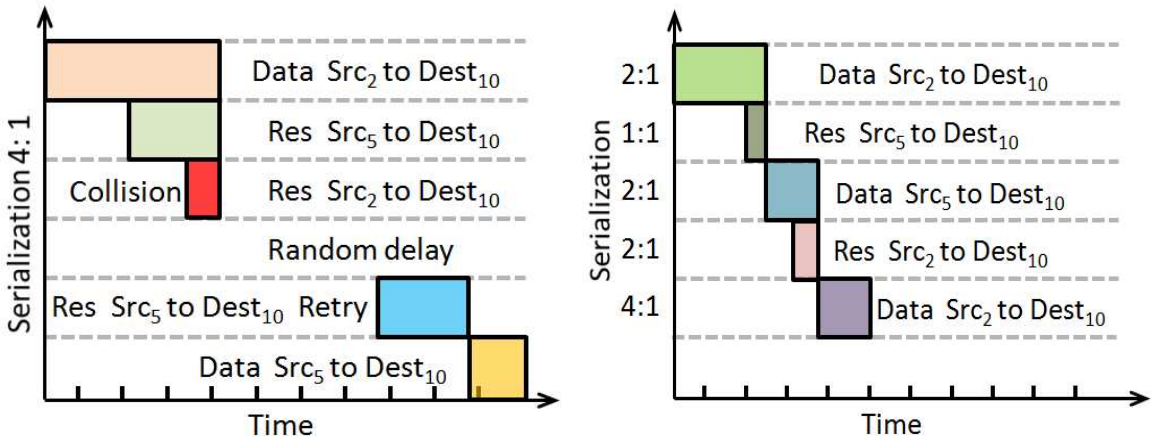


Figure 76 Scenarios for reservation channel collision (a) reservation process with FSNPI collision (b) reservation process after adjusting serialization degree

8.4.7 PCR SIZE SYNTHESIS

In this step, we perform post processing of the combined MEST /MRST and router mapping to develop PCR regions. The root nodes in the MEST that include multiple and multi-depth branches are considered for integration into a PCR region with the nearest gateway interface router as shown in Figure 75(d). More specifically, PCR regions cover nodes that are directly connected to the root nodes with *connection strength* lower than the links between the root nodes. For inter-PCR transfers, we set a size threshold M_{th} such that messages less than the size of M_{th} transverse electrical links, while messages that exceed the threshold size travel to gateway interface routers and utilize FSNPI links. Such a scheme ensures that small message size transfers do not encounter unnecessary E/O and O/E conversion delays, which would make their transfer over FSNPI links less advantageous than over electrical links.

Table 14 MiBench [172] **applications for application categories and number of processors**

Industrial	***Consumer***	***Office***	***Networking***	***Security***
basicmath[8]	jpeg [9]	ghostscript[12]	dijkstra[12]	blowfish[9]
bitcount[11]	lame [6]	rsynth[11]	patricia[14]	rijndael[8]
qsort [8]	mad [7]	stringsearch[11]		sha [9]
susan [9]	tiff2bw[8]			
	tiff2rgba[9]			

8.4.8 CONFLICT ANALYSIS AND RESOLUTION

To reduce collision probability within FSNPI channels, this final step attempts to minimize interference between the various FSNPI communication transactions. Our implementation uses multiple pipelined FSNPI links with separate reservation and data transfer channels, so that the reservation process can proceed while data transmission is in progress. Thus, more than one source core can attempt to reserve the same destination core, resulting in reservation collision

(i.e., interference in modulated photonic links) at the destination node. This collision can produce erroneous data bits.

Such collision can however be detected using parity bits. In our architecture, transaction interference is avoided by managing link bandwidth via modulation of the serialization degree. Transactions from a source router (connected to the initiating core) to the sink router (connected to the target core) along each FSNPI path are evaluated based on detailed communication schedules along a time-axis. In case of any conflicts between two transactions, we serialize these transactions such that both transactions can traverse the same router without interfering with each other. Figure 76 summarizes this process. The channel reservation time is represented by the yellow colored horizontal bar. In the normal case when there is no collision, the reservation proceeds in parallel to data transmission. Once the reservation phase is complete the next data transaction can begin. Figure 76 (a) depicts a collision scenario with the red colored bar, where two reservation requests arrive in parallel with the first transaction's data transmission. This situation requires a reservation retry for the conflicting nodes after a specified retransmission delay thus increasing latency. To eliminate this collision latency, our conflict analysis and resolution step utilizes serialization to modulate communication bandwidth such that multiple streams can coexist without collision. Figure 76 (b) demonstrates how serialization can eliminate retransmission delays due to collision, thereby achieving overall lower transmission latency (at the cost of a slight increase in area and power due to serialization circuitry).

8.5 EXPERIMENTS

8.5.1 APPLICATIONS

We synthesized application-specific hybrid 3D-NoC architectures for five MiBench [173] application benchmark categories: *(i)* Automotive and Industrial Control, *(ii)* Consumer, *(iii)* Office Automation *(iv)* Networking, and *(v)* Security. Applications across these categories generally possess different communication characteristics. As the MiBench benchmarks are written for a single processor, we created our own multithreaded implementations of these benchmarks using Linux *pthreads*. We then generated multi-application workloads that combined multiple MiBench applications executing in parallel, with execution priority assigned to each application in case of any contentions arising during accesses to memories or during task scheduling. We generated instruction and communication traces for the benchmarks via the Shade simulator [172].

Table 15 Communication Synthesis GA Parameter Ranges

Synthesis Parameters	***Range low***	***Range high***
Source Processor ID	1	$m \times n \times l$
Destination Processor ID	1	$m \times n \times l$
Generation Index	1	20
Number of Data Packets	1	4096

Table 14 presents the various application categories, the 17 applications and their corresponding number of threads that we implemented. We created 5 multi-application workloads, corresponding to all applications available in each category, e.g., for the automotive and industrial control multi-application workload, we included parallel implementation of *(i)* basicmath, *(ii)* bitcount, *(iii)* qsort and *(iv)* susan benchmarks. In addition to MiBench

benchmarks, we also evaluated our *3D-HELIX* framework with *NAS* [136] and *PARSEC* [137] application benchmark workloads. The Princeton Application Repository for Shared-Memory Computers (*PARSEC*) benchmark suite is composed of several multithreaded programs that represent next-generation shared-memory programs for CMPs. We considered the following benchmarks: *(i)* blackscholes, *(ii)* bodytrack, *(iii)* canneal, *(iv)* dedup, *(v)* facesim, *(vii)* ferret, *(viii)* fluidanimate, *(ix)*freqmine, *(x)* streamcluster, *(xii)* swaptions, *(xiii)* vips, *(xiv)* x264. *NAS* benchmarks are derived from computational fluid dynamics (CFD) applications. We considered the following benchmarks: *(i)* Embarrassingly Parallel *(ii)* Conjugate Gradient, *(iii)* Multi-Grid, *(iv)* Fourier Transform, *(v)* Integer Sort, *(vi)* Lower-Upper Gauss, *(vii)* Block Tri-diagonal, *(viii)* Scalar Penta.

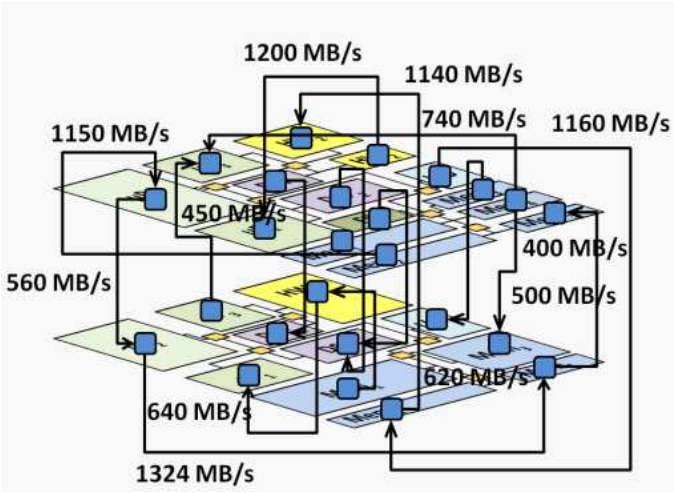
8.5.2 EXPERIMENTAL SETUP

Our core mapping GA *chromosome* consists of parameters with ranges as defined in Table 15. Based on our empirical analysis, we chose an initial population size that included 3200 randomly generated application mappings. We evaluated a range of crossover mutation and probabilities and ultimately utilized probability values of *0.56* and *0.27* respectively. The best fitness value *chromosome* in each iteration that resulted in minimum power consumption while meeting performance constraints was cached to prevent being overwritten by a non-dominated solution chromosome. As the GA is a stochastic search algorithm, it is not possible to formally specify convergence criteria based on optimality therefore we terminated our GA when the best solution quality did not change over a predefined number of iterations (set to a value of 7500).

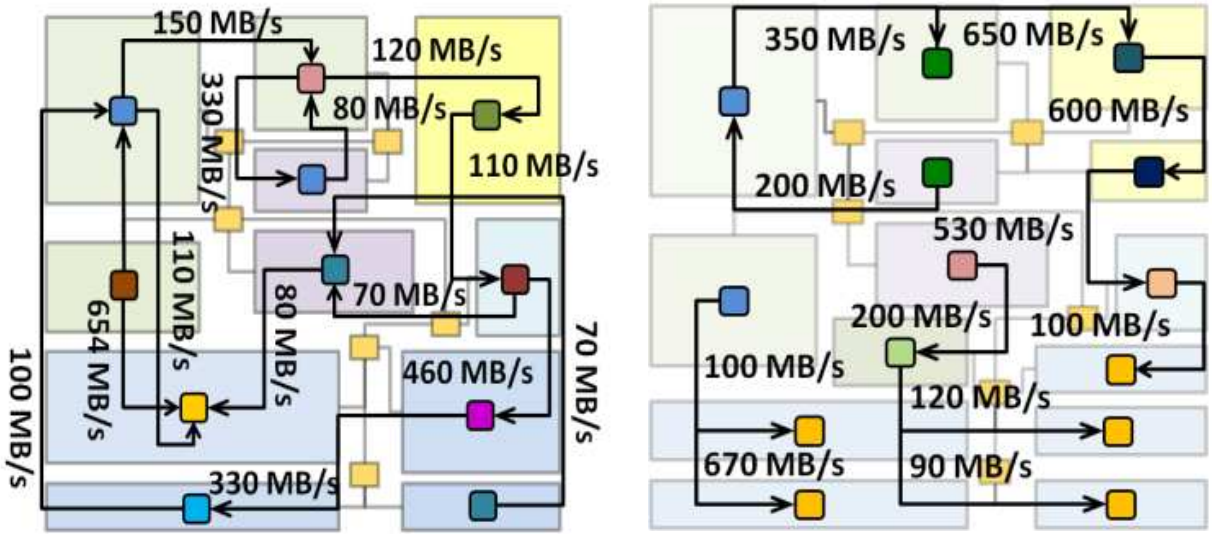
Figure 77(a) shows the enhanced communication trace graph (CTG) of three office applications running in parallel, which is generated after running the GA algorithm and performing core to

die mapping. Note the initial link selection that allocates some flows to electrical links and others to FSNPI links. Figure 77 (b) shows the floorplanning solution for this CTG graph for layer 1 and Figure 77 (c) shows the floorplanning solution for layer 2, where each application is depicted by a separate color. During floorplanning, we set weighed communication power constant $\propto$ and link distance constant β values at 0.5 each based on empirical analysis. We utilized a public domain GeoSteiner3.1 Steiner Tree solver [174] to generate the MEST and MRST networks and utilized the *lp_solve* optimizer [175] to solve the Mixed Integer Linear Programming (MILP). We set weight values for $\propto$ to *0.38* and β to *0.62*, during MEST generation for electrical links and MRST generation for FSNPI links. We created clusters utilizing a *0.44* connection strength threshold to optimize electrical intra-cluster short distance communication and inter-cluster transmission using FSNPI links. Based on our analysis, we set a normalized M_{th} threshold of *39* bytes in PCR regions such that communication messages less than the size of M_{th} transverse through electrical links and the messages that exceed the size of the threshold travel through FSNPI links. We combined the various components of our *3D-HELIX* framework using a python scripting interface [176]. The static and dynamic power consumption of electrical routers as well as the power consumption for optimally sized repeated Cu wires was obtained from a modified version of the Orion 2.0 simulator [141]. Our synthesis process targeted the *18 nm* node technology and utilized a *400* mm^2 SoC die area. The delay of an optimally repeated and sized electrical (Cu) wire at 18 nm was assumed to be *32 ps/mm*. The intrinsic speed of a MQW is practically limited by the driver electronics and the well-known quantum-confined Stark effect working at sub-picosecond time scales. We modeled a $1\mu m$ thick *5V* modulator with $10\times10\mu m^2$ area and capacitance of *11 fF*, calculating the per cycle electrical energy of the device as *140fJ* which is in line with prior estimates [9]. Our implementation assumed modulator driver delay of

9.5 ps, modulator delay of *3.1 ps*, photo detector delay of 0.*22 ps* and receiver delay of *4.9 ps* [36]. Our hybrid NoC was modeled at the cycle accurate granularity by extensively modifying an in-house cycle accurate SystemC-based [158] NoC simulator derived from the Noxim [133] simulator.

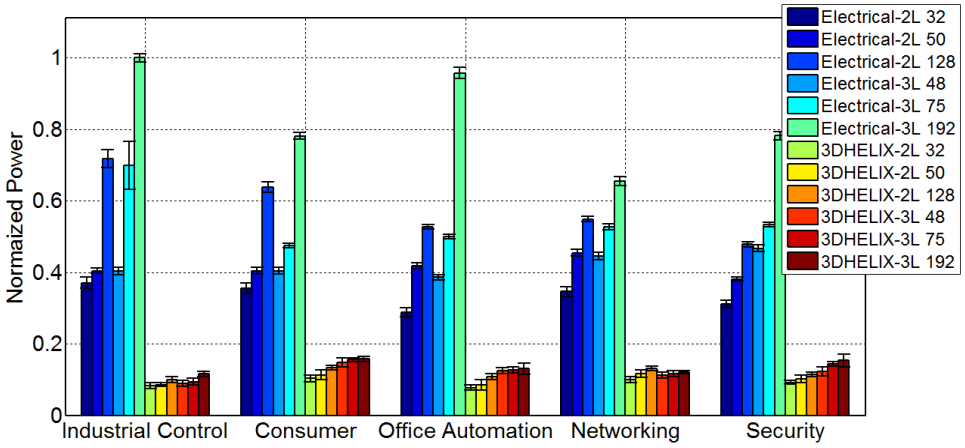


(a)

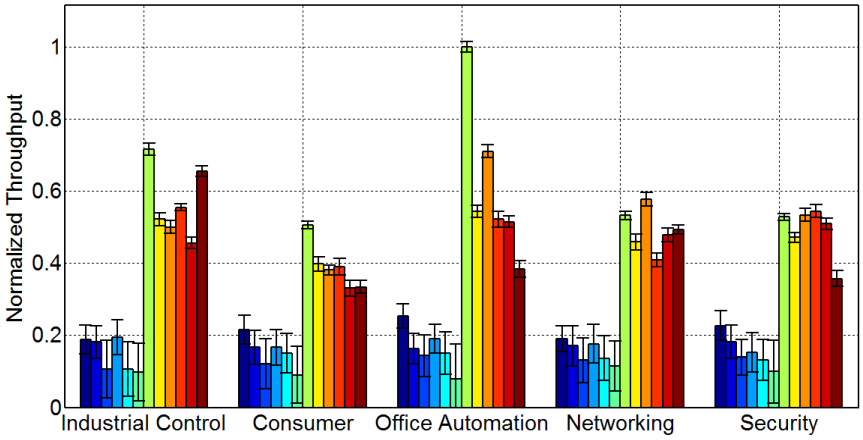


(b) **(c)**

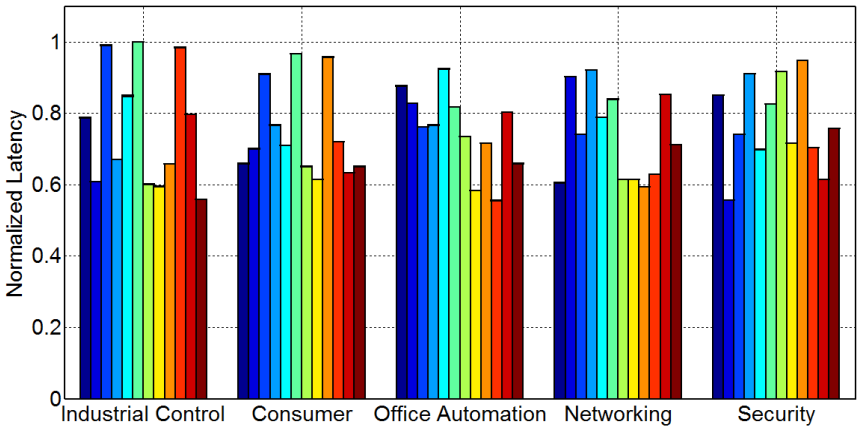
Figure 77 Communication trace graph for multiple parallel applications (a) inter layer communication graph (b) layer 1 communication graph (c) layer 2 communication graph.



(a)

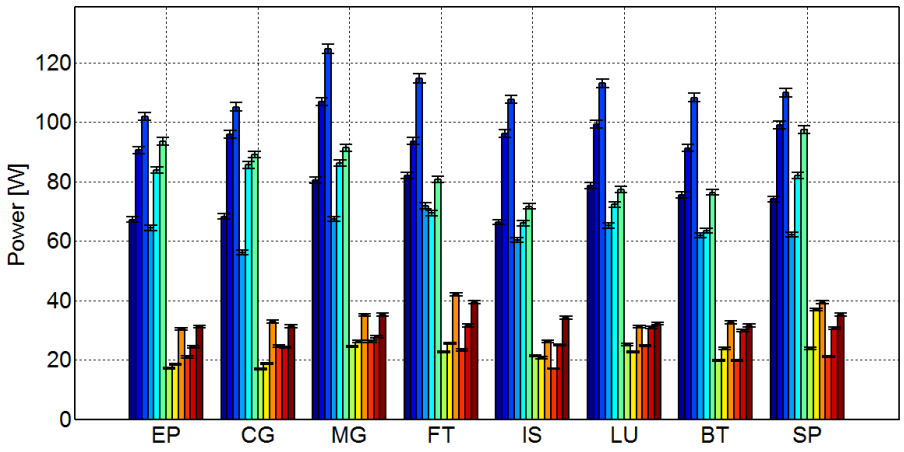


(b)

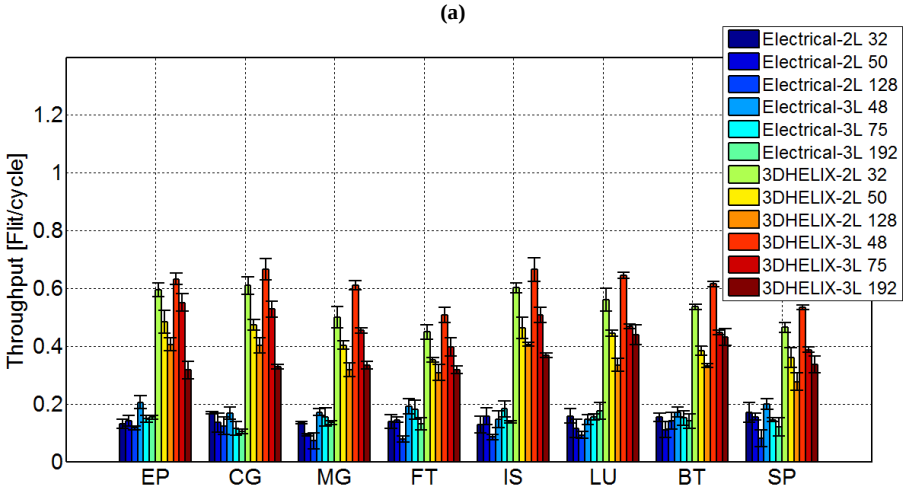


(c)

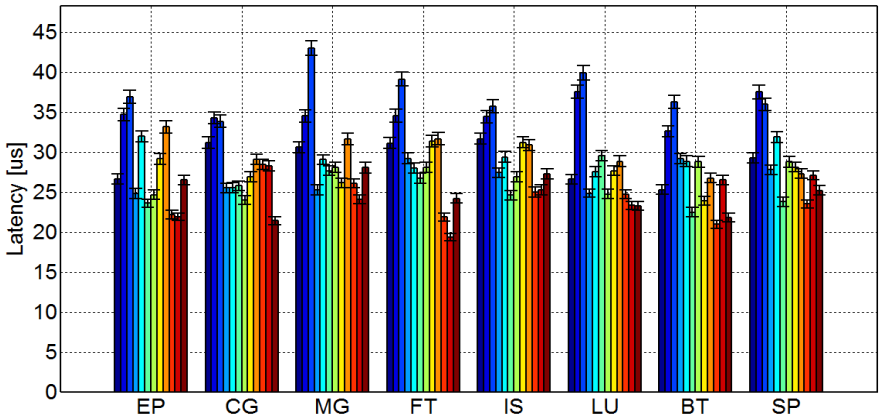
Figure 78 MiBench [173] synthesis results (a) power (b) throughput (c) latency



(a)

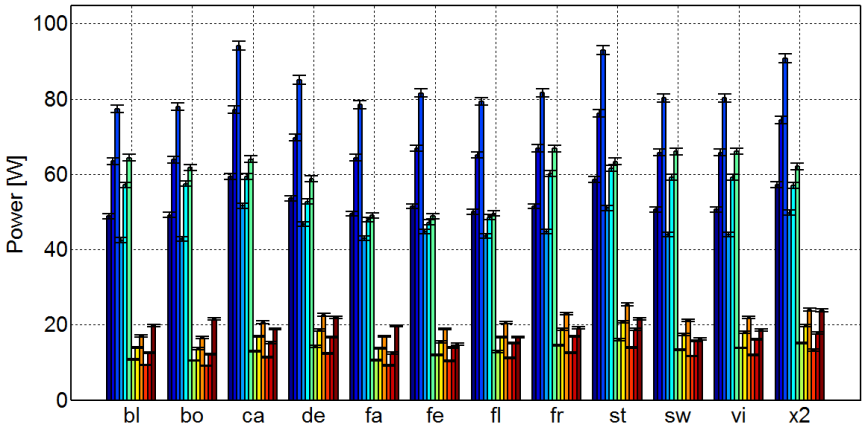


(b)

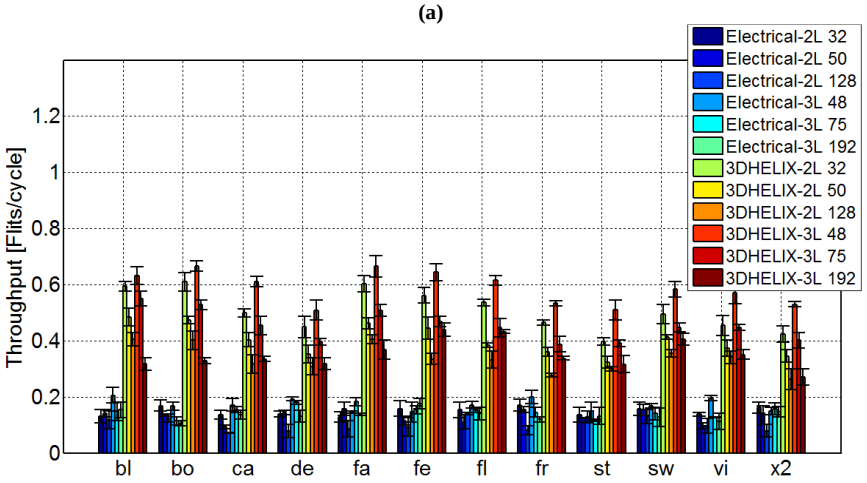


(c)

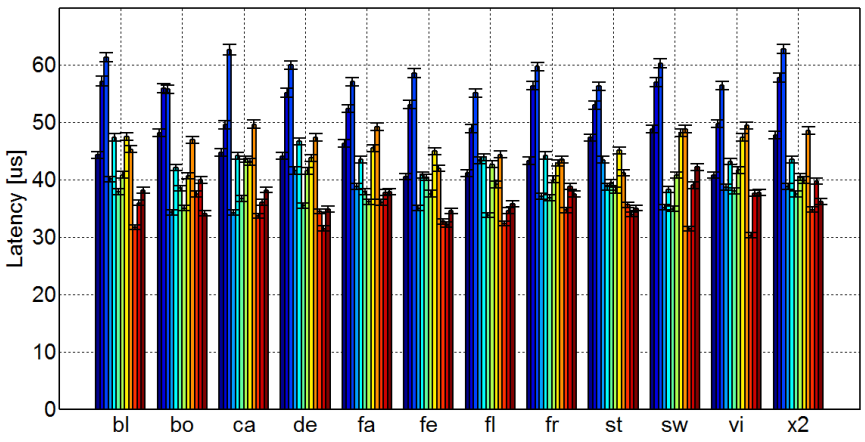
Figure 79 *NAS* [136] synthesis results (a) power (b) throughput (c) tatency



(a)



(b)



(c)

Figure 80 *PARSEC* [137] synthesis results (a) power (b) throughput (c) latency

8.5.3 EXPERIMENTAL RESULTS

In this section, we present experimental results obtained by our *3D-HELIX* framework for the 5 MiBench multi-application workloads [173] and the *NAS* [136] and *PARSEC* [137] benchmarks. To compare the quality of the synthesized solution, we considered synthesized 3D electrical NoCs generated by using a subset of steps from *3D-HELIX* relevant to electrical NoC synthesis (i.e., the photonic link design components were not used). We synthesized hybrid nanophotonic-electric 3D NoCs with small, medium and large sizes defined by (number of cores-number of layers): *32-2L, 50-2L, 128-2L, 48-3L, 75-3L* and *192-3L*; and compared the results against electrical NoCs of the same sizes. Figure 78 shows the results for the MiBench multi-application workloads across the various NoC sizes. Our *3D-HELIX* synthesis framework provides on average *3.66×, 4.11×, 4.97×, 3.53×, 4.30×, 6.16×* reduction in power for *32-2L, 50-2L, 128-2L, 48-3L, 75-3L and 192-3L* NoC platform sizes respectively compared to synthesized application-specific electrical 3D NoCs of the same sizes. *The results indicate the significant potential of including free-space photonic links into 3D-ICs.* The improvement obtained is a result of *(i)* congestion reduction in the electrical links due to offloading of a large portion of the global communication to FSNPI links; *(ii)* reduction in electrical link and buffer switching activity; *(iii)* shorter link lengths and hop counts; and *(vi)* smaller buffer resources, compared to the synthesized 3D electrical application-specific NoC. Our synthesis framework is able to achieve a significant reduction in the number of gateway interface routers through clustering and dual level router mapping, with as few as 37% gateway interface routers compared to the router count before clustering. The dual level router mapping step also added paths enabling inter-cluster long distance global communication using 2-hop FSNPI links to minimize power. We also observed that our conflict analysis and resolution step reduced MQW

modulator and detector counts by approximately 50% by intelligent management of serialization degrees. Figure 79 shows the results for the *NAS* benchmarks [136] (Embarrassingly Parallel (EP), Conjugate Gradient (CG), Multi-Grid (MG), Fourier Transform (FT), Integer Sort (IS), Lower-Upper Gauss (LU), Block Tri-diagonal (BT), Scalar Penta (SP)). For *NAS* [136] workloads, our *3D-HELIX* synthesis framework provides on average *3.46×, 4.01×, 3.28×, 2.86×, 2.71× and 2.50×* reduction in power for *32-2L, 50-2L, 128-2L, 48-3L, 75-3L* and *192-3L* NoC platform sizes respectively, compared to synthesized application-specific electrical 3D NoCs of the same sizes.

Figure 80 shows the results for the *PARSEC* benchmarks [137] (*blackscholes (bl), bodytrack (bo), canneal (ca), dedup (de), facesim (fa), ferret (fe), fluidanimate (fl), freqmine (fr), streamcluster (st), swaptions (sw), vips (vi), x264 (x2)*). For *PARSEC* workloads [137], our *3D-HELIX* synthesis framework provides on average *4.02×, 4.02×, 4.02×, 4.02×, 3.64×* and *3.09×* reduction in power for *32-2L, 50-2L, 128-2L, 48-3L, 75-3L* and *192-3L* NoC platform sizes respectively, compared to synthesized application-specific electrical 3D NoCs of the same sizes.

Our *3D-HELIX* synthesis framework was able to achieve a viable solution for all application workloads and platform complexities that we evaluated. Each stage in our synthesis framework worked seamlessly, complementing each other to balance conflicting requirements to solve the nontrivial problem of synthesizing application-specific hybrid free-space nanophotonic-electric 3D NoC fabrics. The *3D-HELIX* framework took around 6.5 to 14 hours synthesis time depending on the platform size and requirements of the problem being solved. We observed that the number of clusters and gateway interface correlates well with each other and by judiciously selecting FSNPI hop counts, our framework minimizes area and the number of modulators and photodetectors required, without violating performance constraints. This

improvement is possible due to the *3D-HELIX* floorplanner placing cores communicating via FSNPI links farther apart and the cores communicating via electrical links closer, achieving two fold benefits by replacing long distance electrical links with more efficient FSNPI links and placing cores closer that communicate with electrical links. The breakdown of normalized power consumption in Figure 81 demonstrates how *3D-HELIX* can effectively improve power consumption in all categories by managing nontrivial trade-offs during the synthesis process, balancing transfers across the electrical and photonic planes to provide superior results than the synthesized application-specific electrical 3D NoCs.

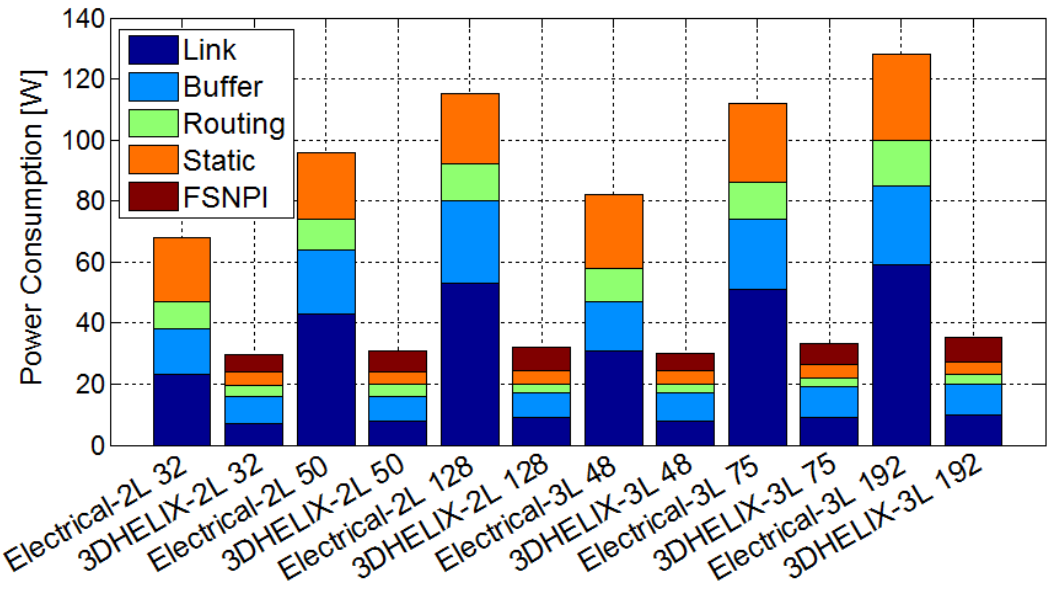


Figure 81 Normalized breakdown of power consumption for 32, 50, 128 core 2-layer and 48, 75, 192 core 3-layer configurations

8.5.4 SUMMARY OF RESULTS

In this chapter, we presented the *3D-HELIX* framework to synthesize heterogeneous application-specific hybrid nanophotonic-electric 3D NoCs for emerging 3D chip multiprocessors. To the best of our knowledge this problem has not been addressed before in any

prior work. Based on our experimental studies, we demonstrate that the proposed techniques in the *3D-HELIX* framework produce a superior hybrid nanophotonic-electric 3D NoC architecture that satisfies all performance requirements for multi-application workloads, while achieving an average from 2.5× to 6× reduction in power for multi-layer small, medium and large sized 3D-NoC based heterogeneous 3D CMP architectures, compared to synthesized application-specific electrical 3D NoCs. We believe that our proposed practical framework will bring hybrid nanophotonic-electric 3D NoCs based on FSNPI and electrical links closer to reality for future heterogeneous 3D CMPs.

9 CONCLUSION AND FUTURE WORK DIRECTIONS

9.1 RESEARCH SUMMARY

As a part of our research, we addressed challenges facing conventional electrical NoCs by proposing novel hybrid electro-photonic NoC architectures and novel synthesis hybrid NoC frameworks for emerging CMPs. Our proposed hybrid electro-photonic NoC architectures are designed for waveguide-based and free-space-based silicon nanophotonics implementations. These architectures are optimized for low-cost, low-power, and low-area overhead, support dynamic reconfiguration to adapt the changing runtime traffic requirements, and have been adapted for both 2D and 3D CMPs. As presented in this thesis, our proposed synthesis frameworks utilize various optimization algorithms such as evolutionary techniques, linear programming, and custom heuristics to perform rapid design space exploration of hybrid electro-photonic (2D and 3D) NoC architectures and trade-off performance and power objectives. Experimental results for our proposed architectures and algorithms indicate a strong motivation to consider them for future CMPs, as they demonstrate several orders of magnitude reduction in power consumption and improvements in network throughput and access latencies, compared to traditional electrical 2D and 3D NoC architectures. Compared to other previously proposed hybrid electro-photonic NoC architectures, our proposed architectures are also shown to have lower photonic area overhead, power consumption, and energy-delay product, while maintaining competitive throughput and latency. Unlike any prior work to date, our synthesis frameworks allow further tuning and customization of our proposed architectures to meet designer-specific goals. Together, the architectural and synthesis framework contributions bring the promise of

silicon nanophotonics in future massively parallel CMPs closer to reality. Following is a quick summary for our key contributions.

Our first contribution is *METEOR*, a novel hybrid nanophotonic-electric NoC architecture that utilizes a low overhead photonic ring waveguide to complement a traditional 2D electrical mesh NoC. *METEOR* includes many novel features such as photonic region of influence (PRI), single write multiple read fast reservation channel and multiple write multiple read (SWMR-MWMR) low power data channels, and serialization for gateway interfaces. Experimental results indicate a strong motivation for considering the *METEOR* hybrid nanophotonic-electric NoCs for future CMPs, demonstrating as much as a 13× reduction in power consumption as well as improved throughput and access latencies, compared to traditional electrical 2D mesh and torus NoC architectures.

An enhancement to the above architecture is proposed to support multiple use-case chip multiprocessor (CMP) applications that require adaptive on-chip communication fabrics to cope with changing use-case performance needs. The proposed *UC-PHOTON* architecture is a novel hybrid nanophotonic-electric NoC communication architecture optimized to cope with the variable bandwidth and latency constraints of multiple use-case applications implemented on CMPs. Detailed experimental results indicate that *UC-PHOTON* can effectively adapt to meet diverse use-case traffic requirements and optimize energy-delay product and power dissipation, with scaling CMP core counts and multiple use-case complexity. For the five multiple use-case applications we explored, *UC-PHOTON* shows up to 46× reduction in power dissipation and up to 170× reduction in energy-delay product compared to traditional electrical NoC fabrics, highlighting the benefits of using the novel communication fabric.

We extended the above work to address scalability issues with 2D ICs and demonstrated a novel multi-layer hybrid nanophotonic-electric NoC fabric called *OPAL* for 3D ICs. Our proposed hybrid nanophotonic-electric 3D NoC architecture combines low cost photonic rings on multiple photonic layers with a 3D mesh NoC in active layers to significantly reduce on-chip communication power dissipation and packet latency. *OPAL* also supports dynamic reconfiguration to adapt to changing runtime traffic requirements, and uncover further opportunities for reduction in power dissipation. Experimental results and comparisons with traditional 2D NoCs, 3D NoCs, and previously proposed hybrid nanophotonic-electric NoCs indicates a strong motivation for considering *OPAL* for future 3D ICs as it can provide orders of magnitude reduction in power dissipation and packet latencies.

As waveguide based hybrid nanophotonic NoCs are constrained due to high thermal tune up power, waveguide crossing losses, we proposed a heterogeneous free space photonics based hybrid nanophotonic-electric NoC architecture with multiple quantum well (MQW) devices and flip chip bonding. The photonic links, using micro-mirrors through free-space can be used to achieve single-hop direct communication links. The proposed architecture includes innovative mechanisms to address free-space collisions and combines single and multi-hop transfers to trade-off performance with power dissipation. We extended this architecture to 3D ICs and the resulting architecture allows for scalable and energy-efficient transfers in 3D ICs.

Any NoC architecture requires optimizations (synthesis) to allow tuning to application-specific characteristics. To date, prior work on automated NoC synthesis has mainly focused on electrical NoCs. For the first time, we proposed a suite of techniques for effectively synthesizing hybrid nanophotonic-electric NoCs with regular topologies by formulating and solving the synthesis problem using four search heuristics: *(i)* Ant Colony Optimization (ACO), *(ii)* Particle

Swarm Optimization (PSO), *(iii)* Genetic Algorithm (GA), and *(iv)* Simulated Annealing (SA). Our experimental results reveal significant promise for the ACO and PSO based heuristics, with PSO achieving an average of *64%* energy-delay improvements over GA and *53%* over SA; and ACO achieving *107%* improvements over GA and *62%* over SA.

Finally, we have designed synthesis frameworks that can synthesize hybrid nanophotonic-electric NoCs with irregular topologies based on free-space photonics. The *HELIX* framework synthesizes application-specific hybrid nanophotonic-electric NoC architectures. This framework is also extended to 3D ICs. The resulting *3D-HELIX* framework synthesizes application-specific hybrid nanophotonic-electric 3D NoC architectures. These frameworks are the first to attempt to optimize hybrid nanophotonic-electric NoC architectures based on free-space photonics.

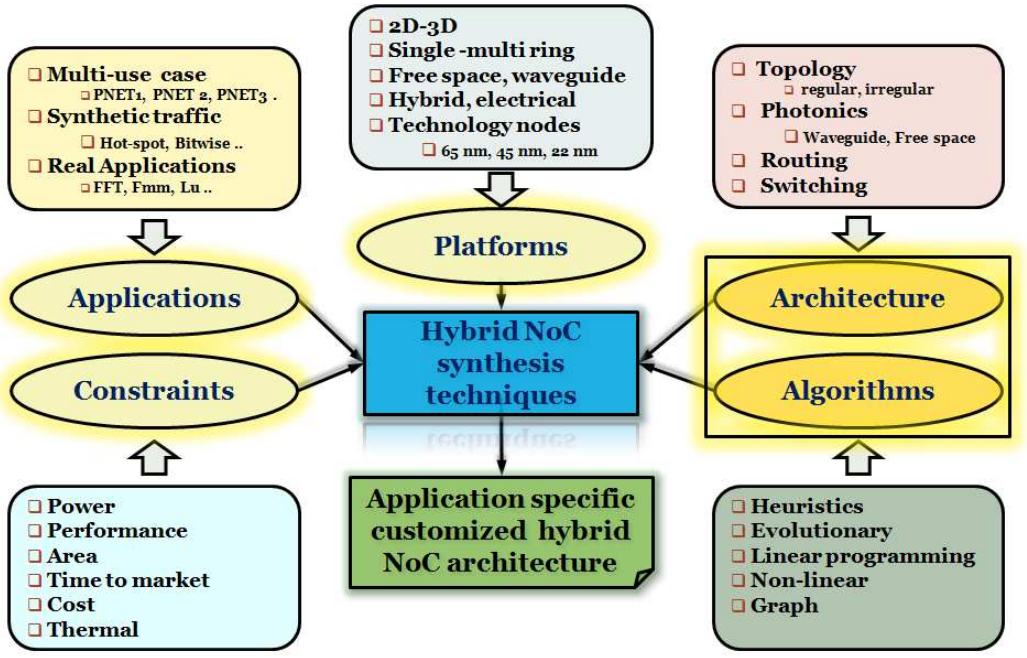


Figure 82 NoC research summary and direction

9.2 FUTURE RESEARCH

Modern digital devices have enriched our life in numerous ways. These modern devices are getting faster, smaller, multifunctional and expected to consume less and less power for every new generation. At the heart of these devices are chip multiprocessors (CMPs) executing multiple use-case applications that require efficient and flexible on-chip communication fabrics to cope with changing use-case performance needs. We have made a case for employing hybrid nanophotonic-electric networks-on-chip (NoC) to address communication challenges in emerging and future CMPs. We further envision the following future work directions:

- Optimizing hybrid NoC's is a daunting task because of the need to traverse through a massive design space. It is highly likely that a communication architecture customized for a single use case may not meet performance requirements for another use case. Thus there is a need to synthesize on-chip communication fabrics to achieve optimal performance for multiple use-case applications. Considering the above unaddressed major challenge, one possible future work extension can be to develop a synthesis framework for application-specific multiple use-case hybrid NoC architectures that combine electrical NoCs with free-space nanophotonic NoCs.
- Another possible direction can be to explore novel architectures based on on-chip free space optical devices to instantiate single-hop or multi-hop direct communication links, that include *(i)* controllable micro-mirrors to utilizing phase change grating to guide the light and *(ii)* on-die VCSELs (vertical cavity surface emitting lasers) to generate light pulses. As VCSELs are directly or indirectly modulated, they have the potential to overcome thermal tune up challenges of silicon microring resonators that are heavily used in waveguide based silicon

photonics. Thermal sensitivity of waveguides with resonators is one of the key reasons to explore free space optical links.

- Hybrid nanophotonic-electrical architecture limitations can possibly be further mitigated by exploring hybrid nanophotonic technologies that combine plasmonic modulators and using photonic-to-plasmonic couplers. Plasmonic modulators tend to be more tolerant to temperature variations compared to silicon microring resonators. Photonic-to-plasmonic couplers can thus enable thermal-invariant modulation using plasmonic modulators and transmit the photonic signals using waveguides.
- Most of our work presented in this thesis was focused on on-chip communication. Extending the work presented in this thesis to off-chip communication can be another possible future direction that can be explored further. Within off-chip hybrid nanophotonic-electrical architecture development work, we can integrate memory controllers and the NoC fabric to evaluate holistic performance and power improvements while also considering external memory devices such as DRAM and emerging PCM devices.
- Another interesting unexplored area so far is developing compilers that support and exploit hybrid nanophotonic-electric on-chip communication architectures. Instruction and data level parallelism enabled by compilers while being cognizant of hybrid nanophotonic-electric NoC architectures can potentially achieve greater performance and power benefits than conventional compiler-unaware approaches.

BIBLIOGRAPHY

[1] S. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, P. Iyer, A. Singh, T. Jacob, S. Jain, S. Venkataraman, Y. Hoskote and N. Borkar., " An 80-Tile 1.28 TFLOPS Network-on-Chip in 65 nm CMOS.," *Proceedings of IEEE International Solid State Circuits Conference*, 2007.

[2] S. Borker, "Design challanges of technology scaling," *IEEE Micro,* vol. 19, no. 4, pp. 23-29, 2009.

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

[4] "ITRS Technology Working Groups, http://public.itrs.net. International Technology Roadmap for Semiconductors (ITRS)," 2007.

[5] D. Pham, S. Asano, M. Bolliger, M. Day, H. Hofstee, C. Johns, J. Kahle, A. Kameyama, J. Keaty, Y. Masubuchi, M. Riley, D. Shippy, D. Stasiak, M. Suzuoki, M. Wang, J. Warnock, S. Weitzel, D. Wendel, T. Yamazaki and K. Yazawa., "The design and implementation of a first-generation CELL processor.," *ISSCC,* pp. 184-185, 2005.

[6] S. Pasricha and N. Dutt, "Trends in Emerging On-Chip Interconnect Technologies," *Proceedings of IPSJ Transactions on System LSI Design Methodology,,* vol. 1, 2008.

[7] J. Owens, W. Dally, R. Ho, D. Jayasimha, S. Keckler and L-S. Peh., "Research challenges for on-chip interconnection networks.," *MICRO,* vol. 27, no. 5, pp. 96-108, 2007.

[8] Corporation Tilera, "Tilera multicore processors.," *http://www.tilera.com/products/processors.*.

[9] D. Miller., "Rationale and challenges for optical interconnects to electronic chips," *Proceedings of J. IEEE,,* June 1994.

[10] R. Ramaswami and K. Sivarajan., Optical Networks: A Practical Perspective., Second ed. Morgan Kaufmann , 2002.

[11] F. Xia, M. Rooks, L. Sekaric and Y. Vlasov., "Ultra-Compact high order ring resonator filters using submicron silicon photonic wires for on-chip optical interconnects.," *Proceedings of Optics Express,* vol. 15, no. 19, pp. 11934-11941, 2007.

[12] W. Green, M. Rooks, L. Sekaric and Vlasov., "Ultra-compact, low RF power, 10 Gb/s silicon mach-zehnder modulator," *Proceedings of International Online Journal of Optics Express (OSA),* vol. 15, no. 25, pp. 17106-17113, May 2007.

[13] C. Schow, F. Doany, O. Liboiron-Ladouceur, C. Baks, D. Kuchta, L. Schares, R. John and J. Kash., "160-gb/s, 16-channel full-duplex, single-chip cmos optical transceiver.," *Proceedings of Optical Fiber Communication Conference,* 2007.

[14] A. Biberman, B. Lee, K. Bergman, P. Dong and M. Lipson., "Demonstration of all-optical multi-wavelength message routing for silicon photonic networks," *Proceedings of Optical Fiber communication/National Fiber Optic Engineers Conference (OFC/NFOEC),* pp. 1-3, Feb 2008.

[15] Y. Vlasov, W. Green and F. Xia., "High-throughput silicon nanophotonic wavelength-insensitive switch for on-chip optical networks.," *Proceedings of Nature Photonics,* vol. 2, no. 4, 2008.

[16] "Lumerical," *http://www.lumerical.com/.*

[17] "Luxtera," *http://www.luxtera.com/.*

[18] C. Gunn, "CMOS photonics for high-speed interconnects," *Proceedings of IEEE Micro (Micro),* vol. 26, no. 2, pp. 58-66, Mar-Apr2006.

[19] "ST-Micro," *http://www.st.com/..*

[20] "Designw," *http://www.designw.com/,* 2011.

[21] W. Haensch., "Is 3D the next big thing in microprocessors?," *Proceedings of International Solid State Circuits Conference (ISSCC),* vol. 12, no. 6, 2007.

[22] S. Bahirat and S. Pasricha, "METEOR: Hybrid Photonic Ring-Mesh Network-on-Chip for Multicore Architectures," *accepted for publication, IEEE Transactions on Embedded Computing Systems (TECS),* 2013.

[23] R. Beausoleil, "Large-scale integrated photonics for high-performance interconnects," *Proceedings of ACM Journal on Emerging Technologies in Computing Systems (JETC) New York, NY, USA,* vol. 7, no. 2, pp. 21-30, June 2011.

[24] B. Koch, A. Fang, O. Cohen and J. Bowers., "Mode-locked silicon evanescent lasers.," *Proceedings of International Online Journal of Optics Express (OE),* vol. 15, no. 18, 2007.

[25] E. DeSouza, M. Nuss, W. Knox and D. Miller, "Wavelength-division multiplexing with femtosecond pulses," *Proceedings of International Online Journal of Optics Express (OSA),* vol. 20, no. 10, pp. 1166-1168, May 1995.

[26] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya and M. Lipson. , "12.5 Gbit/s Carrier-Injection-Based Silicon Microring Silicon Modulators.," *Proceedings of Optical Express,* vol. 15, no. 2, 2007.

[27] A. Gupta, S. Levitan, L. Selavo and D. Chiarulli., "High-speed optoelectronics receivers in SiGe.," in *957-960,* Aug 2004.

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

[29] M. Haurylau, G. Chen, H. Chen, J. Zhang, N. Nelson, D. Albonesi, E. Friedman and P. Fauchet., "On-chip optical interconnect roadmap: challenges and critical directions.," *Proceedings of IEEE Journal of Selected Topics in Quantum Electronics (JSTQE), IEEE,* vol. 12, no. 6, pp. 1699-1705, Nov-Dec 2006.

[30] R. Nair, T. Gu, K. W. Goossen, F. Kiamilev, and M. W. Haney, "Demonstration of chip-scale optical interconnects based on the integration of polymer waveguides and multiple quantum well modulators on silicon," *Proc. IEEE Photonics 2011 Conference, Arlington, VA, USA,* pp. 9-13, October 2011.

[31] L. Chrostowski , B. Faraji , W. Hofmann , R. Shau , M. Ortsiefer and M. C. Amann, "40 GHz bandwidth and 64 GHz resonance frequency in injection-locked 1.55 µm VCSELs," *Proc. Int. Semicond. Laser Conf,* pp. 117-118, 2006.

[32] K. Bernstein, P. Andry, J. Cann, P. Emma, D. Greenberg, W. Haensch, M. Ignatowski, S. Koester, J. Magerlein, R. Puri and A. Young., "Interconnects in the third dimension: design challenges for 3D ICs," *Proceedings of Design Automation Conference (DAC) ACM/IEEE,* pp. 562-567, June 2007.

[33] R. Patti., "Three-dimensional integrated circuits and the future of system-on-chip designs.," *IEEE* , vol. 94, no. 6, 2006.

[34] S. Pasricha, "Exploring Serial Vertical Interconnects for 3D ICs," *IEEE/ACM Design Automation Conference (DAC), in Wormhole Networks,* pp. 519-525, Jul 2009.

[35] S. Bahirat and S. Pasricha, "3D-HELIX: Design and Synthesis of Hybrid Nanophotonic Application-Specific 3D Network-On-Chip Architectures," in *(invited paper) High-*

Performance and Embedded Architectures and Compilers, (HiPEAC), Vienna, Jan. 2014.

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

[37] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang and A. Choudhary., "Firefly: illuminating future network-on-chip with nanophotonics.," *ISCA,* p. 429440, 2009.

[38] A. Shacham, K. Bergman and L. Carloni., "The case for low-power photonic networks on chip.," *DAC,* pp. 132-135, 2007.

[39] L. Zheng, A. Mickelson, L. Shang, M. Vachharajani, D. Filipovic, W. Park and Y. Sun., "Spectrum: A hybrid nanophotonicelectric onchip network.," *Proceedings of Design Automation Conference (DAC),* 2009.

[40] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. Beausoleil and J. Ahn., "Corona: System implications of emerging nanophotonic technology.," *ISCA,* 2008.

[41] R. Morris, A. Kodi, "Power-efficient and high-performance multi-level hybrid nanophotonic interconnect for Multicores," *Proc. International Symposium on Networks-on-Chip (NOCS),* pp. 207-214, 2010.

[42] 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," *Proc. of ISCA,* p. 2010, 94–105.

[43] A. Abousamra, R.Melhem, R., A. Jones, "Two-hop Free-space based optical interconnects for chip multiprocessors," *Proc Fifth IEEE/ACM International Symposium on Networks on Chip (NoCS),* pp. 89-96, 2011.

[44] S. Bahirat and S. Pasricha., "Exploring hybrid photonic networks-on-chip for emerging chip multiprocessors.," in *Proceedings of 7th IEEE/ACM international conference on Hardware/software codesign and system synthesis (CODES+ISSS),* New York, NY, USA, 2009.

[45] S. Murali, P. Meloni, F. Angiolini, D. Atienza, S. Carta, L. Benini, G. De Micheli and L. Raffo, "Designing application-specific networks on chips with floorplan information," *Proc. ICCAD,* pp. 355-362, 2006.

[46] K. Srinivasan, K. S. Chatha, "A low complexity heuristic for design of custom network-on-chip architectures," *Proc. Design, Automation, and Test in Europe (DATE)*, pp. 1-6, 2006.

[47] K. Chatha, K. Srinivasan, and G. Konjevod, "Automated Techniques for Synthesis of Application-Specific Network-on-Chip Architectures," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, pp. 1425-1438, Aug. 2008.

[48] S. Kwon, S. Pasricha, C. Jeonghun, "POSEIDON: A framework for application-specific network-on-chip synthesis for heterogeneous chip multiprocessors," *Proc. International Symposium on Quality Electronic Design (ISQED)*, pp. 182-188, 2011.

[49] S. Murali, G. De Micheli, "Bandwidth-constrained mapping of cores onto NoC architectures," *Proceedings Design, Automation and Test in Europe Conference and Exhibition*, vol. 2, pp. 896-901, Feb. 2004.

[50] G. Ascia, V. Catania, and M. Palesi, "Multi-objective mapping for mesh-based noc architectures," *Proc. International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, pp. 182-187, 2004.

[51] J. Hu, R. Marculescu, "Energy-aware mapping for tile-based noc architectures under performance constraints," *Proc. Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 233-239, 2003.

[52] N. Kapadia, S. Pasricha, "VISION: a framework for voltage island aware synthesis of interconnection networks-on-chip," *Proc. Great Lakes Symposium on VLSI*, p. 2011, 31-36.

[53] S. Bahirat and S. Pasricha, "A Framework for Application-specific Hybrid Photonic Network-on-Chip Synthesis," *Integration, the VLSI Journal*, Under review.

[54] A. Abdelbar, and S. Hedetniemi., "Approximating MAPs for belief networks in NP-hard and other theorems," *Artificial Intelligence*, vol. 102, pp. 21-38.

[55] S. Bahirat, S. Pasricha, "A particle swarm optimization approach for synthesizing application-specific hybrid photonic networks-on-chip," *Proc. International Symposium on Quality Electronic Design (ISQED)*, pp. 78-83, 2012.

[56] S. Kirkpatrick, C. Gelatt Jr., M. Vecchi,, "Optimization by simulated annealing," *Science*, pp. 671-680, 1983.

[57] P. Mazumder, Genetic algorithms for VLSI design, layout and test automation, Prentice-Hall, 1999.

[58] U. Ogras and R. Marculescu., "Its a small world after all: NoC performance optimization via long-range link insertion," *Proceedings of IEEE Transactions on Very Large Scale Integration Systems (VLSI),* vol. 14, no. 7, pp. 693-706, 2011.

[59] A. Kumar, L. Peh, P. Kundu and N. Jha, "Express virtual channels: towards the ideal interconnection fabric," in *ISCA*, 2007.

[60] K. Goossens, J. Dielissen and A. Radulescu., "The thereal network on nhip: noncepts, architectures, and implementations, proceedings of IEEE design and test of computers.," in *Proceedings of IEEE J. Design and Test of Computers (MDT), 414-421*, Sept 2005.

[61] K. Banerjee and A. Mehrotra., "A power-optimal repeater insertion methodology for global interconnects in nanometer designs.," *Proceedings of IEEE Trans. Electron Devices,* vol. 49, no. 11, pp. 2001-2007, Nov 2002.

[62] H. Zhang, V. George and J. Rabaey., "Low-swing on-chip signaling techniques: effectiveness and robustness.," *Proceedings of IEEE Transactions on Very Large Scale Integration Systems (VLSI),* vol. 8, no. 3, 2000.

[63] P. Larsson-Edefors, D. Eckerbert, H. Eriksson and L. Svensson., "Dual threshold voltage circuits in the presence of resistive interconnects.," *Proceedings of IEEE Computer Society Annual Symposium on VLSI (ISVLSI),* pp. 225-230, 2003.

[64] R. Bashirullah, W. Liu, and R. K. Cavin, III., "Current-mode signaling in deep submicrometer global interconnects.," *Proceedings of IEEE Trans.Very Large Scale Integration (VLSI) Systems,* vol. 11, no. 3, pp. 406-417, June 2003.

[65] V.Wang, G. Pei and E. Kan., "Pulsed wave interconnect," *Proceedings of IEEE Trans. Very Large Scale Integration Systems (VLSI),* vol. 12, no. 5, pp. 453-463, 2004.

[66] H. Kaul, D. Sylvester., "A novel buffer circuit for energy efficient signaling in dual-VDD systems.," in *Proceedings of 15th ACM Great Lakes Symposium on VLSI (GLSVLSI), 462-467*, IEEE/ACM Press New York, NY, USA, 2005.

[67] M. Dresselhaus, G. Dresselhaus, P. Avouris and R. Smalley., Carbon nanotubes: synthesis, structure, properties, and applications., Springer (ISBN 978-3-540-41086-7), 2001.

[68] F. Kreup, A. Graham, M. Liebau, G. Duesberg, R. Seidel and E. Unger., " Carbon nanotubes for interconnect applications.," in *IEDM*, Dec 2004.

[69] N. Srivastava and V. Banerjee., "Performance analysis of carbon nanotube interconnects for VLSI applications.," *ICCAD*, 2005.

[70] S. Pasricha and N. Dutt., "ORB: An On-chip optical ring bus communication architecture for multi-processor systems-onchip.," *ASPDAC*, 2008.

[71] M-C. Chang, S Frank, S. Tam, J. Cong and G. Reinman, "RF interconnects for communications on-chip," in *Proceedings of the 2008 international symposium on Physical design (ISPD), ACM Press*, New York, NY, USA, March 2008, 78-83.

[72] D. Zhao and Y. Wang. 2008., "SD-MAC: Design and synthesis of a hardware-efficient collision-free QoS-aware MAC protocol for wireless network-on-chip.," *Proceedings of IEEE Transactions on Computers*, vol. 57, no. 9, pp. 1230-1245, 2008.

[73] F. Li, C. Nicopoulos, T. Richardson, Y. Xie, V. Narayanan and M. Kandemir, "Design and Management of 3D Chip Multiprocessors Using Network-in-Memory," *ISCA*, pp. 130-141, 2008.

[74] B. Feero and P. Pande., "Performance evaluation for three-dimensional networks-on-chip," in *Proceedings of the IEEE Computer Society Annual Symposium on VLSI (ISVLSI), IEEE*, 305-310, March 2007.

[75] Z. Li, X. Hong, Q. Zhou, J. Bian,H. Yang, H. Hannah V. Pitchumani, "Efficient thermal-oriented 3D floorplanning and thermal via planning for two-stacked-die integration," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 11, no. 2, pp. 325--345, April 2006.

[76] E. Wong, S. K. Lim, "3D Floorplanning with Thermal Vias," *Proc. DATE*, pp. 1-6, 2006.

[77] P. Zhou, Y. Ma, Z. Li, S. Li, H. Zhou, X. Hong, Q. Zhou, "3D-STAF: scalable temperature and leakage aware floorplanning for three-dimensional integrated circuits," *International Conference on Computer-Aided Design. ICCAD. IEEE/ACM*, pp. 590 - 597, 2007.

[78] Y. Liu, Y. Ma, E. Kursun, G. Reinman, "Fine grain 3D integration for microarchitecture design through cube packing exploration," *Oct. 2007*, pp. 259-266, International Conference on Computer Design. ICCD.

[79] D. Park, S. Eachempati, R. Das, A.K. Mishra, X. Yuan, N. Vijaykrishnan, C. Das, "MIRA: A Multi-layered On-Chip Interconnect Router Architecture," *35th International Symposium on Computer Architecture. ISCA*, pp. 251 - 261, June 2008.

[80] K. Puttaswamy, G. H. Loh, "Implementing caches in a 3D technology for high performance processors," *Proc. ICCD*, 2005.

[81] I. Loi, F. Angiolini, L. Benini., "Supporting vertical links for 3D networks-on-chip: toward an automated design and analysis flow," *Proceedings of the 2nd international conference*

on Nano-Networks (Nano-Net), vol. 15, pp. 1-5, 2007.

[82] J. Goodman, F. Leonberger, K. Sun-Yuan and R. Athale., "Optical interconnects for VLSI systems.," *Proceedings of IEEE Optical interconnections for VLSI systems (PROC),* vol. 72, no. 7, pp. 850-866, July 1984.

[83] D. Chiarulli, S. Levitan, R. Melhem, M. Bidnurkar, R. Ditmore, G. Gravenstreter, Z. Guo, J. Qao and C. Teza., "Optoelectronic buses for high performance computing," *Proceedings of the IEEE (SLIP),* vol. 82, no. 11, pp. 1701-1710, Nov 1994.

[84] J. Ha and T. Pinkston., "Speed Demon: cache coherence on an optical multichannel interconnect architecture.," *Proceedings of Journal of parallel and distributed computing (JPDC), Elsevier,* vol. 41, no. 1, pp. 78-91, Feb 1997.

[85] E. Carrera and R. Bianchini., "OPTNET: A cost-effective optical network for multiprocessors.," in *Proceedings of the 12th international conference on Supercomputing (ICS),* ACM Press New York, NY, USA, 401-408, June 2008.

[86] A. Kodi and A. Louri, "Rapid: Reconfigurable and scalable all-photonic in-104 interconnect for distributed shared memory multiprocessors.," *Proceedings of Journal of Light-wave Technology,* vol. 22, pp. 2101-2110, 2004.

[87] C. Kochar, A. Kodi and A. Louri., "Nd-Rapid: a multidimensional scalable fault-tolerant optoelectronic interconnection for high performance computing systems.," *Proceedings of Journal of Optical Networking,* vol. 6, no. 5, 2007.

[88] M. Tan, P. Rosenberg, Y. Jong-Souk, M. McLaren, S. Mathai, T. Morris, K. Pei, J. Straznicky, N. Jouppi and S. Wang., "A high-speed optical multi-drop bus for computer interconnections.," *Proceedings of 16th IEEE Symposium on High Performance Interconnects,* p. 2008, 3-10.

[89] G. Chen, H. Chen, M. Haurylau, N. Nelson, D. Albonesi, M. Philippe, P. Fauchet, E. Friedman, and G. Eby., "Predictions of CMOS compatible on-chip optical interconnect," in *Proceedings of the 2005 international workshop on System level interconnect prediction (SLIP), ACM Press*, New York, NY, USA, 13-20, June 2005.

[90] J. Collet, F. Caignet, F. Sellaye and D. Litaize., "Performance constraints for onchip optical interconnects.," *Proceedings of IEEE J. Selected Topics in Quantum Electronics (JSTQE),) IEEE,* vol. 9, no. 2, pp. 425-432, July 2003.

[91] G. Tosik, F. Gaffiot, Z. Lisik, I. O'Connor and F. Tissafi-Drissi., "Power dissipation in optical and metallic clock distribution networks in new VLSI technologies.," *Proceedings*

of IEEE Electronics Letters, vol. 40, no. 3, pp. 198-200, 2004.

[92] M. Kobrinsky, B. Block, J. Zheng, B. Barnett, E. Mohammed, M. Reshotko, F. Robertson, S. List and I. Young., "On-Chip Optical Interconnects.," *Intel Technology,* vol. 8, no. 2, pp. 129-142, 2004.

[93] I. OConnor., "Optical solutions for system-level interconnect," *Proceedings of the 6th international workshop on System level interconnect prediction (SLIP),* Vols. ACM Press New York, NY, USA, pp. 79-88, 2004.

[94] A. Pappu and A. Apsel., "Analysis of intrachip electrical and optical fanout.," *Proceedings of Applied Optics,* vol. 44, no. 30, p. 63616372, 2005.

[95] A. Biberman, K. Preston, G. Hendry, N. Sherwood-Droz, J. Chan, J. Levy, M. Lipson and K. Bergman., "Photonic network on chip architectures using multilayer deposited silicon materials for high-performance chip multiprocessors," *Proceedings of ACM Journal on Emerging Technologies in Computing Systems (JETC) ACM Press New York, NY, USA,* vol. 2, no. 7, pp. 7-15, June 2011.

[96] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, M. Lipson and M. Lipson., "12.5 Gbit/s carrier-injection-based silicon microring silicon modulators.," *Proceedings of Optics Express,* vol. 15, no. 2:22, p. 430436, 2007.

[97] S. Sahni, X. Luo, J. Liu, Y. Xie, and E. Yablonovitch, "Junction field-effect-transistor-based germanium photodetector on silicon-on-insulator," *Proc. Optics Letters,* vol. 33, pp. 1138-1140, May 2008.

[98] A. Okyay, D. Kuzum, D. S. Latif, D. Miller, et al., "Silicon Germanium CMOS Optoelectronic Switching Device: Bringing Light to Latch," *IEEE Transactions on Electron Devices,* vol. 54, no. 12, pp. 3252 - 3259, Dec. 2007.

[99] A. Biberman, B. Lee, K. Bergman, A. Turner-Foster, M. Lipson, M. Foster, A. Gaeta, "First Demonstration of On-Chip Wavelength Multicasting," *Conference on Optical Fiber Communication OFC,* pp. 1-3, March 2009.

[100] Z. Li, M. Mohamed, X. Chen, H. Zhou, A. Mickelson, L. Shang and M. Vachharajani, "Iris: A hybrid nanophotonic network design for high-performance and low-power on-chip communication," *Proceedings of ACM Journal on Emerging Technologies in Computing Systems (JETC),* vol. 7, no. 2, p. 6, (2011.

[101] N. Krman, M. Krman, R. Dokania, J. Martnez, A. Apsel, M. Watkins, D. Albonesi., "Leveraging Optical Technology in Future Bus-based Chip Multiprocessors. (MICRO),,"

2006.

[102] A. Shacham, K. Bergman and L. Carloni, "Photonic networks-on chip for future generations of chip multiprocessors.," *Proceedings of IEEE Transactions on Computers,* vol. 59, no. 9, pp. 1246-1260, 2008.

[103] M. Cianchetti, J. Kerekes and D. Albonesi., "Phastlane: A rapid transit optical routing network.," *Proceedings of the 36th annual international symposium on Computer architecture (ISCA), ACM Press New York, NY, USA,* vol. 3, no. 37, pp. 441-450, June 2009.

[104] X. Zhang and A. Louri., "A multilayer nanophotonic interconnection network for on-chip many-core communications.," *Proceedings of Design Automation Conference (DAC),* pp. 156-161, 2011.

[105] Y. Kao and H. Chao, "BLOCON: A Bufferless Photonic Clos network-on-chip architecture.," in *Proceedings of ACM/IEEE International Symposium on Networks-on-Chip (NOCS),* IEEE/ACM Press New York, NY, USA, 2011.

[106] R. Morris and A. Kodi., "Exploring the design of 64- and 256-core power efficient nanophotonic interconnect.," *Proceedings of IEEE Journal of Selected Topics In Quantum Electronics,* 2010.

[107] A. Joshi, J. Orcutt, A. Khilo, B. Moss, C. Holzwarth, M. Popovic, L. Hanqing, H. Smith, J. Hoyt, F. Kartner, R. Ram, V. Stojanovic and K. Asanovic., "Building many core processor-to-dram networks with monolithic silicon photonics," *Proceedings of 16th Annual Symposium on High-Performance Interconnects,* pp. 21-30, Aug 2008.

[108] S. Koohi and S. Hessabi., "Power efficient nanophotonic on-chip network for future large scale multiprocessor architectures.," in *Proceedings of IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH),* 2011.

[109] A. Hansson, K. Goossens, A. Rădulescu, "A unified approach to constrained mapping and routing on network-on-chip architectures," *International conference on Hardware/software codesign and system synthesis,* pp. 75- 80, 2005.

[110] J. Hu, R. Marculescu, "Exploiting the Routing Flexibility for Energy/Performance Aware Mapping of Regular NoC Architectures," *Design, Automation and Test in Europe Conference and Exhibition,* pp. 688-693, 2003.

[111] A. Hansson , K. Goossens, "Trade-Offs in the Configuration of a Network on Chip for Multiple Use-Cases," *International Symposium on Networks-on-Chip,* pp. 233 - 242, May

2007.

[112] I. Loi, F. Angiolini, L. Benini, "Synthesis of low-overhead configurable source routing tables for network interfaces," *Design, Automation and Test in Europe Conference and Exhibition*, pp. 262 - 267, Apr. 2009.

[113] M. Al Faruque, T. Ebi, J. Henkel, "Configurable links for runtime adaptive on-chip communication," *Design, Automation and Test in Europe Conference and Exhibition, DATE* , pp. 256 - 261, April 2009.

[114] M. Li, Q. Zeng, W. Jone, "DyXY - a proximity congestion-aware deadlock-free dynamic routing method for network on chip," *Design Automation Conference*, pp. 849 - 852, 2006.

[115] K. Lahiri, A. Raghunathan, G. Lakshminarayana, S. Dey, "Design of high-performance system-on-chips using communication architecture tuners," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* , vol. 23, no. 5, pp. 620 - 636, May 2004.

[116] T. Richardson, C. Nicopoulos, V. Narayanan et al., "A hybrid SoC interconnect with dynamic TDMA-based transaction-less buses and on-chip networks," *International Conference on VLSI Design and Embedded Systems and Design*, pp. 657-664, Jan. 2006.

[117] S. Pandey, T. Murgan, M. Glesner, "Energy Conscious Simultaneous Voltage Scaling and On-Chip Communication Bus Synthesis," *International Conference on Very Large Scale Integration, IFIP*, pp. 296-301, Oct. 2006.

[118] S. Pasricha, N. Dutt, F. Kurdahi, "Dynamically reconfigurable on-chip communication architectures for multi use-case chip multiprocessor applications," *Proceedings of the 2009 Asia and South Pacific Design Automation Conference*, pp. 25-30, Jan 2009.

[119] S. Murali, M. Coenen, A. Radulescu, K. Goossens, G. De Micheli, "A methodology for mapping multiple use-cases on to networks on chip," *Proceedings Design, Automation and Test in Europe, DATE,* vol. 1, pp. 1-6, 2006.

[120] S. Murali, M. Coenen, A. Radulescu, K. Goossens, G. De Micheli, "Mapping and configuration methods for multi-use-case networks on chips," *Asia and South Pacific Conference on Design Automation,* vol. 1, pp. 146-151, Jan. 2006.

[121] A. Hansson, M. Coenen, K. Goossens, "Undisrupted Quality-of-Service during Reconfiguration of Multiple Applications in Networks on Chip," *Design, Automation and Test in Europe Conference and Exhibition*, pp. 1-6, 2007..

[122] P. Koonath and B. Jalali, "Multilayer 3-d photonics in silicon," *Opt. Express,* vol. 15, no. 20, p. 12686–12691, 2007.

[123] H. Wassel, D. Dai, L. Theogarajan, J. Dionne, M. Tiwari, J. Valamehr, F. Chong and T. Sherwood, "Opportunities and challenges of using plasmonic components in nanophotonic architectures.," *Proceedings of IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS),* 2012.

[124] Y. Xie, J. Xu, J. Xu and J. Zhang., "Elimination of cross-talk in silicon-on-insulator waveguide crossings with optimized angle.," *Proceedings of Optical Engineering,* vol. 50, no. 6, pp. 064601-064604, 2011.

[125] Z.H. Ai-Awwami, M.S. Obaidat and M. Al-Mulhem, "ZOMA: a preemptive deadlock recovery mechanism for fully adaptive routing in wormhole networks," in *International Conference on Computer Networks and Mobile Computing (ICCNMC),* 2001.

[126] R. Dokania and A. Apsel., "Analysis of challenges for on-chip optical interconnects," in *Proceedings of 19th ACM Great Lakes symposium on VLSI (GLSVLSI),* ACM Press New York, NY, USA, 275-280, May 2009.

[127] A. Morgenshtein, I. Cidon, A. Kolodny and R. Ginosar., "Comparative analysis of serial Vs parallel links.," *SSOC,* 2004.

[128] M. Ghoneima, Y. Ismail, M. Khellah, J. Tschanz and V. De., "Serial-link bus: A low-power on-chip bus architecture.," *Proceedings of IEEE Transactions on Circuits and Systems I (TCSI),* vol. 56, no. 9, pp. 2020-2032, Dec 2005.

[129] S. Kimura, T. Hayakawa, T. Horiyama, M. Nakanishi and K. Watanabe., "An On-Chip high speed serial communication method based on independent ring oscillators. (ISSCC)," 2003.

[130] I. Wey, L. Chang, Y. Chen, S. Chang and A. Wu., "A 2Gb/s high-speed scalable shift-register based on-chip serial communication design for SoC applications," *Proceeding of Circuits and Systems (ISCAS),* pp. 468-469, June 2005.

[131] M. Saneei, A. Afzali-Kusha and M. Pedram., "Two high performance and low power serial communication interfaces for on-chip interconnects.," *CJECE,* 2008.

[132] "Nirgam simulator," *http://nirgam.ecs.soton.ac.uk/.*

[133] "Noxim simulator," *http://noxim.sourceforge.net/.*

[134] S. Adya, and I. Markov., "Fixed-outline floorplanning: enabling hierarchical design," *J. IEEE Very Large Scale Integration (VLSI) Systems,* vol. 11, pp. 1120-1135, Dec 2003.

[135] S. Woo and M. Ohara., "The SPLASH-2 programs: characterization and methodological considerations.," *Proceedings of of the International Symposium on Computer Architecture (ISCA),* no. 2436, 1995.

[136] H. Jin, M. Frumkin and J. Yan., "The OpenMP Implementation of NAS Parallel Benchmarks and Its Performance, NASA Ames Research Center. Technical Report NAS-99-011," Edition [Online]. Available: citeseer.ist.psu.edu/408248.html, 1999.

[137] C. Bienia, S. Kumar, J. Singh and K. Li., "The PARSEC benchmark buite: characterization and architectural implications.," in *Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques (PACT),* ACM Press New York, NY, USA, 72-81, Oct 2008.

[138] T. Barwicz, H. Byun, F. Gan, C.W. Holzwarth, M.A. Popovic, P.T. Rakich, M.R. Watts, E.P. Ippen, F.X. Krtner, H.I. Smith, J.S. Orcutt, R.J. Ram, V. Stojanovic, O. Olubuyide, J.L. Hoyt, S. Spector, M. Geis, M. Grein, T. Lyszczarz and J.U. Yoon., "Silicon photonics for compact, energy-efficient interconnects.," *J. Optical Networking,* vol. 6, no. 1, pp. 63-73, Jan 2007.

[139] J. Cunningham, S. Ivan, Z. Xuezhe, P. Thierry, M. Attila, L. Ying, T. Hiren, L. Guoliang, Y. Jin, R. Kannan and A. Krishnamoorthy., "Highly-efficient thermally-tuned resonant optical filters.," *Proceedings of International Online Journal of Optics Express (OSA),* vol. 18, no. 18, pp. 19055-19063, Sept 2010.

[140] I-W. Hsieh, X. Chen, J. Dadap, N. Panoiu, J. Osgood, S. McNab and Y. Vlasov., "Ultrafast-Pulse Self-Phase Modulation and Third-Order Dispersion in Si Photonic Wire-Waveguides," *Proceedings of International Online Journal of Optics Express (OE),* vol. 14, no. 25, pp. 12380-12387, Dec 2006.

[141] A. Kahng, B. Li, L. Peh and K. Samadi., "ORION 2.0: A Fast and Accurate NoC Power and Area Model for Early-Stage Design Space Exploration. (DATE)," in *423-428,* 2009.

[142] S. Kawanishi, H. Takara, K. Uchiyama, I. Shake and K. Mori., "3 Tbit/s (160 Gbit/s19 channel) optical TDM and WDM transmission experiment," *Proceedings of Electronic Letters,* vol. 35, no. 10, p. 82627, 1999.

[143] M. Watts., "Ultralow power silicon microdisk modulators and switches.," *Proceedings of 5th Annual Conference on Group IV Photonics,* 2008.

[144] Y. Vlasov and S. McNab., "Losses in single-mode silicon-on-insulator strip waveguides and bends.," *Proceedings of Optical Express,* vol. 12, no. 8, 2004.

[145] D. Ding and D. Pan, "OIL: A nano-photonics optical interconnect library for a new photonic networks-on-chip architecture.," in *Proceedings of the 11th international workshop on System level interconnect prediction (SLIP),* ACM Press New York, NY, USA, 11-18, June 2009.

[146] J. Ahn, M. Fiorentino, R.G. Beausoleil, N. Binkert, A. Davis, D. Fattal, N.P. Jouppi, M. McLaren, C.M. Santori, R.S. Schreiber, S.M. Spillane, D. Vantrease and Q. Xu., "Devices and architectures for photonic chip-scale integration.," *J. Applied Physics A Materials Science and Processing,* vol. 95, no. 4, pp. 989-997, Dec 2009.

[147] P. Koka, M. McCracken, H. Schwetman, X. Zheng, R. Ho and A. Krishnamoorthy., "Silicon-photonic network architectures for scalable, power-efficient multi-chip systems.," *ISCA,* pp. 117-128, 2010.

[148] A. Suga and K. Matsunami, "Introducing the FR 500 embedded microprocessor," *IEEE MICRO,* vol. 20, p. 21–27, 2000.

[149] J Cornish, "Balanced energy optimization," *Proc ISLPED,* 2004.

[150] T. Hattori, Y. Yoshida, K. Hayase,T. Hayashi, O. Nishii, Y. Yasu, A. Hasegawa, M.Takada, H. Mizuno, K. Uchiyama, T. Odaka, J. Shirako, M. Mase, K. Kimura, H. Kasahara, "An 8640 MIPS SoC with Independent Power-Off Control of 8 CPUs and 8 RAMs by An Automatic Parallelizing Compiler," *Proc. ISSCC,* 2008.

[151] W. Dally and B. Towles., "Route packets, not wires: on-chip interconnection networks.," in *Proceedings of Design Automation Conference (DAC), IEEE,* 684-689, May 2001.

[152] M. Taylor, J. Kim, J. Miller, D. Wentzlaff, F. Ghodrat, B. Greenwald, H. Hoffmann, P. Johnson, J. Lee, W. Lee, A. Ma, A. Saraf, M. Seneski, N. Shnidman, V. Strumpen, M. Frank, S. Amarasinghe, A. Agarwal, "The Raw Microprocessor: A Computational Fabric for Software Circuits and General-Purpose Programs," *IEEE Micro,* vol. 22, no. 2, pp. 25-35, 2002.

[153] R. Beausoleil, J. Ahn, N. Binkert, A. Davis, D. Fattal, M. Fiorentino, N. Jouppi, M. McLaren, C. Santori, R. Schreiber, S. Spillane, D. Vantrease, Q. Xu, "A Nanophotonic Interconnect for High-Performance Many-Core Computation," *Hot Interconnects,* pp. 182-189, 2008.

[154] A. Covell, F. Whyte, "Digital Convergence: How the Merging of Computers, Communications and Multimedia is Transforming Our Lives," *Aegis Publishing Group,* 1999.

[155] L. Schares, A. Kash, F. Doany, C. Schow, C.L. et al., "Terabus: Terabit/Second-Class Card-Level Optical Interconnect Technologies," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 12, no. 5, pp. 1032-1044, Sep/Oct 2006.

[156] W. Bogaerts, P. Dumon , D. Thourhout, D. Taillaert, P. Jaenen, J. Wouters, S. Beckx, S. V. Wiaux, V. R. Baets, "Compact Wavelength-Selective Functions in Silicon-on-Insulator Photonic Wires," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 12, no. 6, pp. 1394 - 1401, Dec. 2006.

[157] J. Rabaey, A. Chandrakasan, B. Nikolic, Digital Integrated Circuits, Prentice Hall, 2002.

[158] "SystemC initiative," *www.systemc.org.*

[159] S. Pasricha, N. Dutt, M. Ben-Romdhane, "Extending the transaction level modeling approach for fast communication architecture exploration," *In Proc. of DAC,* pp. 113-118, 2004.

[160] I. Artundo, W. Heirman, M. Loperena, C. Debaes, J. Van Campenhout, H. Thienpont, "Low-Power Reconfigurable Network Architecture for On-Chip Photonic Interconnects," *IEEE Symposium on High Performance Interconnects, HOTI,* pp. 163-169, Aug. 2009.

[161] K. Srinivasan, K. S. Chatha and G. Konjevod, "Linear Programming Based Techniques for Synthesis of Network-on-Chip Architectures," *In Proceedings of ICCD,* pp. 422-429, October 2004.

[162] J. Chan, S. Parameswaran, "NoCOUT: NoC topology generation with mixed packet-switched and point-to-point networks," *Proc. Asia and South Pacific Design. Automation Conference (ASP-DAC),* pp. 265-270, 2008.

[163] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Proc. Journal of Chemical Physics,* vol. 21, no. 6, pp. 1087-1092, 1953.

[164] R. Eberhart, J. Kenned,, "A new optimizer using particle swarm theory," *Proc. International Symposium on Micromachine and Human Science,* pp. 39-43, Nagoya, Japan. 1995..

[165] C. Teo, Y. Foo, S. Chien, A Low, B. Venkatesh, A. You, "Optimal placement of

wavelength converters in WDM networks using particle swarm optimizer," *Proc. IEEE International Conference on Communications,* pp. 1669-1673, 2004.

[166] P. Tawdross, A. Konig, "Investigation of particle swarm optimization for dynamic reconfiguration of field-programmable analog circuits," *Proc. Fifth International Conference on Hybrid Intelligent Systems,* p. 6, 2005.

[167] M. Dorigo, V. Maniezzo, A. Colorni, "Ant system: optimization by a colony of cooperating agents," *Proc. IEEE Transactions on Systems, Man, and Cybernetics--Part B,* vol. 1, p. 26, 1996.

[168] V. Černý, "A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm.," *Proc. Journal of Optimization Theory and Applications,* vol. 45, pp. 41-51, 1985.

[169] M. Lundy, A. Mees, "Convergence of an Annealing Algorithm," *Math. Prog.,* vol. 34, pp. 111-124, 1986.

[170] D. Goldberg, Genetic algorithms in search, optimization and machine learning, Boston, MA: Kluwer Academic Publishers, 1989.

[171] G. Tian, R. Nair, M. Haney, "Integrated free-space optical interconnects: All optical communications on- and off-chip," *Optical Interconnects Conference, IEEE,* pp. 74-75, 2012.

[172] B. Cmelik and D. Keppel, "Shade: A fast instruction-set simulator for execution profiling," *SIGMETRICS '94: Proceedings of the 1994 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems,* pp. 128-137, 1994.

[173] M. Guthaus, J. Ringenberg, D. Ernst, T. Austin, T. Mudge, and R. Brown, "MiBench: A Free, Commercially Representative Embedded Benchmark Suite," *IEEE 4th Annual Workshop on Workload Characterization (WWC-4),* December 2001.

[174] "Geosteiner," *http://www.diku.dk/geosteiner/.*

[175] "lpsolve," *http://lpsolve.sourceforge.net/5.5/.*

[176] "Python," *http://www.python.org/.*

[177] D. Miller, "Energy consumption in optical modulators for interconnects," *Opt. Express,* pp. A293-A308, 2012.

[178] S. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, A. Singh, T. Jacob, S. Jain, V. Erraguntla, C. Roberts, Y. Hoskote, N. Borkar and S. Borkar., "An 80-tile sub-100-w teraflops processor in 65-nm cmos.," *Proceedings of IEEE J. of Solid-State Circuits,*, vol. 43, no. 1, p. 2941, Jan 2008.

[179] M-C. Chang, J. Cong, A. Kaplan, M. Naik, G. Reinman, E. Socher and S-W. Tam., "OPTNET: A cost-effective optical network for multiprocessors.," in *Proceedings of 14th International Symposium on High Performance Computer Architecture (HPCA), IEEE,* 191-202, Feb 2008.

[180] M. Cianchetti and D. Albonesi., "A low-latency, high-throughput on-chip optical router architecture for future chip multiprocessors.," *Proceedings of ACM Journal on Emerging Technologies in Computing Systems (JETC), ACM Press New York, NY, USA,* vol. 7, no. 2, pp. 1-20, July 2011.

[181] D. Costantini, H. Limberger, R. Salathe, C. Muller and S. Vasiliev., "Tunable loss filter based on metal coated long period grating.," in *Proceedings of European Conference on Optical Communication (ECOC), IEEE, 391-392,* Sept 1999.

[182] R. Dobkin, A. Morgenshtein, A. Kolodny and R. Ginosar., "Parallel vs. serial on-chip communication.," in *Proceedings of the 10th international workshop on System level interconnect prediction (SLIP),* ACM Press New York, NY, USA, 43-50, June 2008.

[183] B. Guha, B. Kyotoku and M Lipson., "CMOS-compatible athermal silicon microring resonators," *Proceedings of International Online Journal of Optics Express (OE),* vol. 18, no. 4, pp. 3487-3493, Feb 2010.

[184] A. Hemani, A. Jantsch, S. Kumar, A. Postula, J. Oberg, M. Millberg and D. Lindqvist., "Network on Chip: an architecture for billion transistor era.," in *Proceeding of the IEEE NorChip Conference (NORCHIP),* IEEE,166-173, Nov 2000.

[185] Y. Hoskote, S. Vangal, A. Singh, N. Borkar and S. Borkar., "A 5-GHz mesh interconnect for a teraflops processor," *Proceedings of IEEE Micro (Micro),* vol. 27, no. 5, pp. 51-61, Sept-Oct 2007.

[186] WorkGroup ITRS, "ITRS Technology Working Groups," in *International Technology Roadmap for Semiconductors (ITRS),* 2007.

[187] Z. Li, M. Mohamed, X. Chen, E. Dudley, K. Meng, L. Shang, A. Mickelson, R. Joseph, M. Vachharajani, B. Schwartz and Y. Sun., " Reliability modeling and management of nanophotonic on-chip networks," *Proceedings of IEEE Transactions on Very Large Scale*

Integration Systems (VLSI), 2011.

[188] 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," *Proceedings of Nature,* vol. 427, no. 615618, 2004.

[189] X Liu, A. Liao, L. Chetrit, Y. Basak, J. Nguyen, H. Rubin and D. Paniccia., "Wavelength division multiplexing based photonic integrated circuits on silicon-on-insulator platform.," *Proceedings of IEEE J. Select. Topics Quantum Electron,* vol. 16, no. 1, p. 2332, 2010.

[190] G. Merrett, B. Al-Hashimi, "Leakage power analysis and comparison of deep submicron logic gates," *PATMOS,* pp. 198-207, 2007.

[191] D. Miller, J. Weiner, D. Chemla., "For a summary of work on quantum well electroabsorption and further discussion of QCSE theory.," *Proceedings of IEEE J. Quantum Electron,* 1986.

[192] C. Nitta, M. Farrens and V. Akella., "Resilient microring resonator based photonic networks.," *MICRO,* pp. 95-104, 2011.

[193] S. Pasricha, S. Bahirat,, "OPAL: A multi-layer hybrid photonic NoC for 3D ICs.," *Proceedings of IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC),* 2011.

[194] S. Pasricha, F. Kurdahi and N. Dutt., "System Level Performance Analysis of Carbon Nanotube Global Interconnects for Emerging Chip Multiprocessors.," *Proceedings of IEEE/ACM NanoArch,,* 2008.

[195] Q. Xu, B. Schmidt, S. Pradhan and M. Lipson., "Micrometre-scale silicon electro-optic modulator.," *Proceedings of Nature Letters,* vol. 435, 2005.

[196] L. Zheng, M. Moustafa, Z. Hongyu, S. Li, M. Rolf, F. Dejan, M. Vachharajani, W. Park and Y. Sun., "Global On-Chip Coordination at Light Speed.," *Proceedings of Test of Computers, ACM Transactions,* vol. 27, no. 4, pp. 54-67, 2010.

[197] S. Koohi, S. Hessabi, "Contention-Free on-Chip Routing of Optical Packets," *ACM/IEEE International Symposium on Networks-on-Chip, NoCS* , pp. 134-143, May 2009.

[198] H. Gu, J. Xu, W. Zhang, "A Low-Power Fat Tree-based Optical Network-on-Chip for Multiprocessor System-on-Chip," *Design, Automation and Test in Europe Conference and Exhibition, DATE,* pp. 3-8, April 2009.

[199] "System Level Assessment of an Optical NoC in an MPSoC Platform," *Design, Automation and Test in Europe Conference and Exhibition, DATE*, pp. 1-6, Apr. 2006.

[200] L. Benini and G. De-Micheli, "Networks on Chip: A new SoC paradigm," *Proc. Computer*, vol. 49, no. 1, pp. 70-71, Jan 2002.

[201] J. Goodman, F. Leonberger, S. -Y. Kung, R. Athale, "Optical interconnects for VLSI systems," *Proc. IEEE*, vol. 26, no. 1, pp. 850-866, July 1984.

[202] S. Murali, P. Meloni, F. Angiolini, D. Atienza, S. Carta, L. Benini, G. D. Micheli, L. Raffo, "Designing application-specific networks on chips with floorplan information," *Proc. International Conference on Computer-Aided Design (ICCAD)*, p. 2006, 355-362.

[203] "Plurality HAL-256," *http://www.plurality.com/products.html*, 2009.

[204] S. Bahirar and S. Pasricha, "UC-PHOTON: A Novel Hybrid Photonic Network-on-Chip for Multiple," *ISQED*, 2010.

[205] S. Bahirat, "Design and Synthesis of Hybrid Nanophotonic NoCs for Future Many-Core Architectures," in *DAC PhD Forum*, Austin, 2013.

[206] M. Oxley, S. Pasricha, A. Maciejewski, H.J. Siegel, J. Apodaca, B. Young, L. Briceno, J. Smith, S. Bahirat and B. Khemka, "Makespan and Energy Robust Stochastic Static Resource Allocation of Bags-of-Tasks to a Heterogeneous Computing System," *(Under review) IEEE Transactions on Parallel and Distributed Systems*, 2014.

[207] D. Young, J. Apodaca, L. Briceno, J. Smith, S. Pasricha, A. Maciejewski, H. Siegel, S. Bahirat, B. Khemka, A. Ramirez and Y. Zou, "Deadline and Energy Constrained Dynamic Resource Allocation in a Heterogeneous Computing Environment," *Journal of Supercomputing*, vol. 63, no. 1, pp. pp 326-347, February 2013.

[208] J. Apodaca, B. Young, Luis Diego Briceno, J. Smith, S. Pasricha, A. Maciejewski, H. Siegel, S. Bahirat, B. Khemka, A. Ramirez and Y. Zou, "Stochastically robust static resource allocation for energy minimization with a makespan constraint in a heterogeneous computing environment," in *(Best Paper Award) IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, Sharm El-Sheikh, 2011.