OPAL: A Multi-Layer Hybrid Photonic NoC for 3D ICs

Sudeep Pasricha, Shirish Bahirat

Colorado State University, Fort Collins, CO
{sudeep, shirish.bahirat}@colostate.edu

Abstract - 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 paper we propose and explore a novel multi-layer hybrid photonic 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.

I Introduction

Rapid improvements in CMOS fabrication technology and a steady rise in application complexity in recent years have led to the emergence of chip multiprocessors (CMPs) as compact and powerful computing paradigms. Emerging CMP designs are integrating in the order of a hundred or more cores on a chip [1]-[3]. To cope with the growing communication demands of these massively multi-core systems, shared bus-based communication fabrics are being replaced by networks-on-chip (NoCs) that offer higher reliability, scalability, and bandwidths [4][42]. In practice, the large number of network interfaces (NIs), routers, links, and buffers in NoCs lead to significant power dissipation, e.g., ~30% in the Intel 80-core teraflop chip [2] and ~40% in the MIT RAW chip [3]. 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 [5]. Thus, radical new approaches are required to overcome the power and performance brick walls facing NoCs in the near future [6].

Of the several different disruptive technologies that are being investigated today, 3D integrated circuits (3D-ICs) with wafer-to-wafer bonding technology is one of the most promising candidates [7][8][45]. 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 [2], IBM [7], and Tezzaron [8] have confirmed the benefits of 3D IC technology.

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, novel interconnect materials need to be explored. *Photonic interconnects* [9] are an extremely promising emerging solution that can replace Cu interconnects and help overcome their latency and power bottlenecks. Photonic interconnect technology can transfer data with much more energy efficiency than Cu interconnects especially over long distances on chip. In addition, the ability of photonic waveguides to carry many information channels simultaneously using wavelength division multiplexing (WDM) increases interconnect bandwidth density significantly, eliminating the need for a large number of wires to achieve bandwidth goals. Photonic interconnects are becoming standard in data centers, and chip-to-chip photonic links have been demonstrated [10]. This trend will naturally bring photonic interconnects into the on-chip stack, particularly as a means to enable high bandwidth and low power data transfers between hundreds of cores in future CMPs. Recent advances in the field of silicon photonics have enabled highly integrated photonic interconnect-based components in CMOS-based ICs [11]-[14].

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 been addressed so far. The question arises: *can hybrid photonic NoCs be viable interconnect fabrics for future 3D ICs?* In this paper, we attempt to answer this question, and propose OPAL, a novel hybrid 3D NoC architecture that combines low cost photonic rings on multiple photonic layers with 3D mesh NoC fabrics in active layers. The photonic paths offload global communication from the electrical network, improving packet latency and reducing communication power dissipation. In addition, OPAL supports dynamic reconfiguration of the electrical and photonic networks. This enables runtime adaptation to changing traffic volumes, which allows network resources to be optimized for even lower 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 several orders of magnitude reduction in power dissipation and average packet latencies.

II. Related Work

Over the last several years, there has been a growing interest in 3D ICs as a means to alleviate the interconnect bottleneck currently facing 2D ICs. 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 [15]-[17]. A few researchers have explored interconnect architectures for 3D ICs such as 3D mesh and stacked mesh NoC topologies [18] and a hybrid bus-NoC topology [19]. Some recent work has looked at decomposing cores (processors [20], NoC routers [21], and on-chip cache [22]) 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 [23].

Recent advances in silicon photonics have led to the development of fabrication technologies to stack optical devices in multiple layers [24] in 3D ICs. Literature abounds in comparisons of the physical properties of on-chip electrical and photonic interconnects [25]-[27] at the circuit level, highlighting the signal speed and power benefits of photonic interconnects. Other work from industry and academia has been focusing on photonic device fabrication, e.g., gigascale modulators [11], photodetectors [12], switches [13], couplers, buffers, waveguides and on-chip wave division multiplexing (WDM) devices [14]. A few recent works have explored the system- level impact of using hybrid electro-photonic interconnect architectures and proposed hybrid Cu-photonic crossbars (Corona [28], Firefly [29]), Clos networks [30], fat-trees [31] and torii [32][33]. However these architectures possess high area and fabrication complexity (e.g. more than a million resonators in Corona [28]). Our recent work [34] explored a simpler WDM-enabled hybrid Cu-photonic architecture with a parallel photonic ring waveguide interfaced to an electrical 2D mesh NoC. While this architecture is more area and cost effective than other hybrid electro-photonic topologies (e.g., ~15× lower photonic layer overhead vs. hybrid photonic torus [32]), it is not scalable for large CMP designs.

In this paper, we propose a hybrid 3D multi-ring/mesh topology to improve performance scalability for emerging CMPs with hundreds of cores. Runtime reconfiguration of the electrical and photonic networks is explored with the goal of significantly reducing communication power dissipation. Several works [35]-[37] have explored runtime electrical NoC adaptation schemes including DVS/DFS, dynamic

978-1-4244-7516-2/11/$26.00 ©2011 IEEE

routing schemes, and dynamic arbitration to adapt to changing runtime traffic needs, and improve performance and power dissipation. However, previous work has not explored runtime reconfiguration in 3D hybrid electro-photonic NoCs.

III. OPAL Overview

A. Photonic Building Blocks

Fig. 1 shows a high level overview of the primary on-chip 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 [38] enable WDM that allows several data streams to coexist in the same waveguide, improving transfer bandwidth. In this work, we assume that wavelengths are allocated to traffic streams using 'multiplexing by core', with each of the n interfacing cores having exclusive access to λ/n wavelengths, where λ is the total number of wavelengths supported. This limits the number of transmitters, but provides substantial power savings.

Microring resonant modulators [11] convert electrical data signals into light, which is propagated through a CMOS-compatible silicon oxide photonic waveguide. The light in the waveguide is eventually coupled into microring filters at the destination that drop the light on photodetectors [39], 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 vital for all microring resonators to be thermally tuned (using thermal heater elements) to maintain their resonance under on-die temperature variations [30].

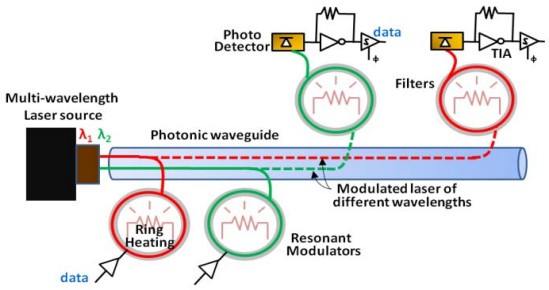


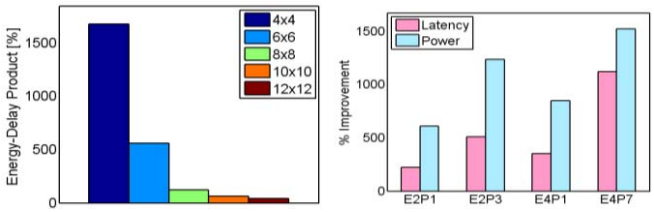
Fig 1: Building blocks of photonic interconnects

To accurately estimate performance and power overhead of on-chip photonic communication, we characterize the delay and power overhead of each of the described photonic components, including the thermal heaters and the laser. As laser power is determined by the magnitude of losses in photonic components, we account for losses due to couplers, resonators, photodetectors, waveguide length and bends, and non-linearity. Section IV.A summarizes the per-component delay, power, and loss characterization considered in this work in detail.

B. 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 *single photonic layer* has the lowest design complexity, but may lack scalability. For instance, if a hybrid electro-photonic torus topology [32] 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 [34] 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. Fig. 2(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 [40] 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.



(a) (b)

Fig 2: 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

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. Fig 2(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; Fig. 3), *(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. Fig. 2(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. We present more comprehensive experimental results in Section IV. In the following sections, we describe our multi-layer hybrid photonic NoC architecture in detail.

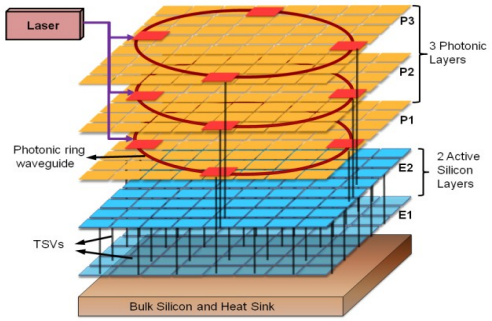


Fig 3: E2P3 OPAL configuration

C. OPAL System Level Architecture

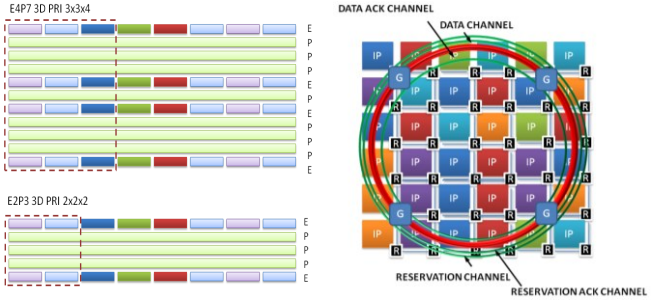
In this work, we propose OPAL, which is a hybrid electro-photonic 3D NoC architecture that employs multiple active layers and multiple photonic layers with photonic ring paths in a stack. The active layers consist of cores interconnected to each other using a 3D electrical mesh NoC. The photonic layers 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.

Fig 3 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 paper 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.

D. 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. Fig 4(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).



(a) (b)

Fig 4: (a) 3D Photonic Region of Influence (3D-PRI) (b) Photonic channels

E. 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.

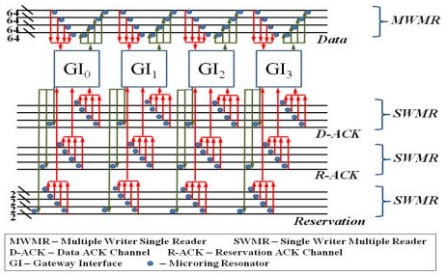


Fig 5: SWMR reservation channels and MWMR data channels

F. 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, as shown in Fig 4 (b) and Fig. 5. 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_i) 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 [28] and Firefly [29]. In OPAL, the number of data channel waveguides is equal to the chosen flit width (i.e., 256). The same wavelength (λ_i) 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.

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 [32]-[34], 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 [28].

G. 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 [47]. 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.

H. 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: ***(i) 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 [41]; ***(ii) 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; ***(iii) 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 to 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.

IV. Experiments

A. 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 [40] were selected (*Cholesky, FFT, Fmm, Lu, Radix, Ocean*), parallelized, and implemented on multiple cores in the simulator model. Fig. 6 shows an example of the SPLASH-2 benchmarks implemented on an 8x8 CMP. Each cell represents a core, with lighter colored cores sending/receiving fewer packets than darker colored cores.

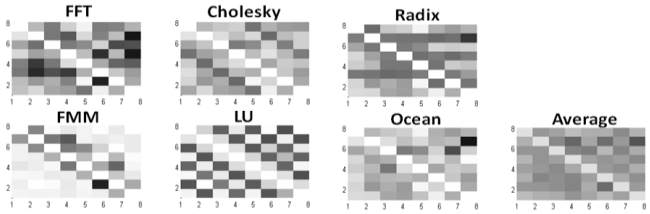


Fig 6: SPLASH-2 implementation traffic maps for 8x8 CMP

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, using delay values from Table 1, 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 [43] and from device fabrication results [44]. Table 1 shows these delays for the 32 nm node. The delay of an optimally repeated and sized electrical (Cu) wire at 32 nm was assumed to be 42ps/mm [9].

TABLE I
Delay and energy consumption for OPAL elements (32nm) *DDE = Data traffic dependent energy, SE = Static energy (clock, leakage), TTE = Thermal tuning energy (20K temperature range)*

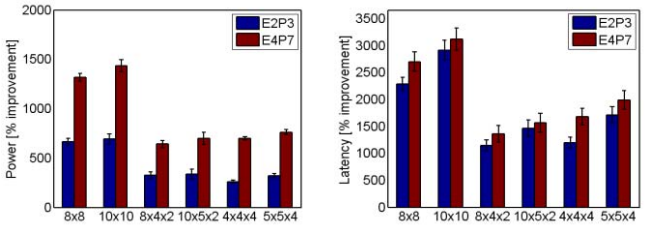
Component	Delay	DDE	SE	TTE
Modulator driver	9.5 ps	20 fJ/bit	5 fJ/bit	16 fJ/bit/heater
Modulator	3.1 ps			
Waveguide	15.4 ps/mm	-	-	-
Photo Detector	0.22 ps	20 fJ/bit	5 fJ/bit	16 fJ/bit/heater
Receiver	4.0 ps			

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 [46] incorporated into our simulator. For calculating power dissipation of the modulator driver and TIA power we used ITRS device projections [6] and standard circuit procedures. Energy dissipation values for the modulators and receivers are summarized in Table 1 [30]. 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).

B. Results

B.1 Comparisons with 2D and 3D Electrical Mesh NoC

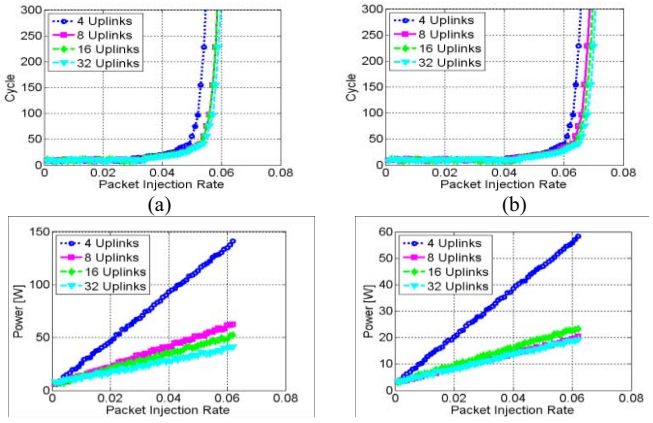
In the first set of experiments, we compared the performance and power characteristics of the E2P3 and E4P7 OPAL configurations, but without enabling any dynamic reconfiguration, with traditional 2D and 3D electrical mesh NoCs. The 2D configurations considered included a 64 core (8×8) and 100 core (10×10) 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.



(a) (b)

Fig 7: percentage improvement for OPAL configurations compared to 2D and 3D electrical mesh NoCs (a) power, (b) average packet latency

Fig 7 (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 8× power reduction and 20× 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.



(a) (b)

(c) (d)

Fig 8: 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

B.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. Fig 8 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.

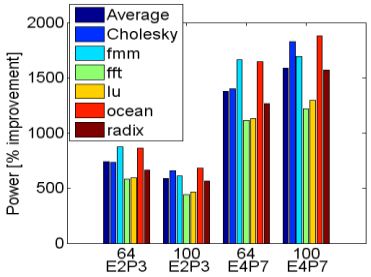


Fig 9: percentage improvement in average power dissipation for E2P3 and E4P7 OPAL configurations, with all runtime adaptations enabled (DVS/DFS, WDM, PRI) relative to baseline case with no runtime adaptation enabled

B.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. As discussed in Section III.H, dynamically adapting resources based on runtime traffic requirements can expose opportunities for power savings. Fig 9 presents results of power savings for the six SPLASH-2 benchmark implementations when the dynamic reconfiguration schemes described in Section III.H (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. It can be seen from Fig 9 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.

B.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 [32], *(ii)* the hybrid Corona architecture [28], and *(iii)* the hybrid Firefly architecture [29]. 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. Fig 10 (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. 10× 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. 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.

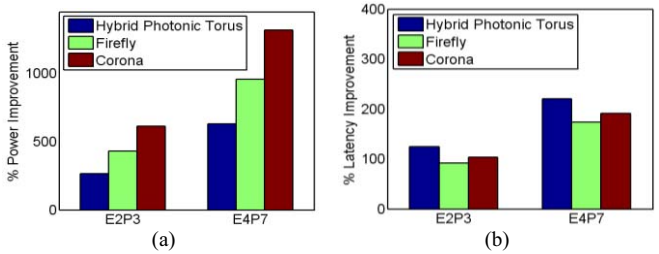


Fig 10: percentage improvement for E2P3 and E4P7 OPAL configurations compared with hybrid photonic torus [32], Corona [28] and Firefly [29] NoCs: (a) power dissipation (b) average packet latency

V. Summary and Conclusions

In this paper, 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. Our future work will explore the thermal design considerations in 3D hybrid electro-photonic NoCs in more detail.

References

[1] Plurality HAL-256, http://www.plurality.com/products.html, 2009.
[2] S. Vangal et al., "An 80-Tile 1.28 TFLOPS Network-on-Chip in 65 nm CMOS," Proc. IEEE Int'l Solid State Circuits Conf., Feb. 2007.
[3] Tilera Corporation. TILE64™ Processor. Product Brief. 2007.
[4] L. Benini and G. De-Micheli, "Networks on Chip: A New SoC Paradigm," Proc. Computer, 49(1):70-71, Jan 2002.
[5] J.D. Owens, et al., "Research challenges for on-chip interconnection networks," Proc. IEEE Micro, 27(5):96–108, 2007.
[6] ITRS Technology Working Groups, http://public.itrs.net. International Technology Roadmap for Semiconductors (ITRS) 2007 Edition.
[7] K. Bernstein, et al., "Interconnects in the Third Dimension: Design Challenges for 3D ICs," Proc. DAC 2007, pp.562-567.
[8] R. S. Patti, "Three-Dimensional Integrated Circuits and the Future of System-on-Chip Designs", Proc IEEE, Vol 94, No. 6, Jun 2006.
[9] M. Haurylau, et al., "On-chip Optical Interconnect Roadmap: Challenges and Critical Directions," IEEE JQE 12(6), Nov 2006.
[10] L. Schares, et al., "Terabus: Terabit/Second-Class Card-Level Optical Interconnect Technologies", IEEE JQE, 12(5), Sep/Oct 2006.
[11] Q. Xu, et al., "12.5 Gbit/s Carrier-Injection-Based Silicon Microring Silicon Modulators," Proc. Optics Express, 15(2):430-436, Jan 2007.
[12] S. Sahni et al. "Junction Field-effect-transistor based Germanium Photodetector on Silicon-on-Insulator", Proc. Opt. Letters, May 2008
[13] A. K. Okyay et al., "Silicon Germanium CMOS Optoelectronic Switching Device: Bringing Light to Latch", IEEE TED, Dec 2007.
[14] A. Biberman et al., "First Demonstration of On-Chip Wavelength Multicasting", In Optical Fiber Comm. Conference, March 2009
[15] Z. Li, et al., "Efficient thermal-oriented 3D floorplanning and thermal via planning for two-stacked-die integration", TODAES 11:2, Apr 2006.
[16] E. Wong, S. K. Lim, "3D Floorplanning with Thermal Vias", Proc. DATE 2006, pp. 1-6.
[17] P. Zhou et al., "3D-STAF: scalable temperature and leakage aware floorplanning for three-dimensional integrated circuits", Proc ICCAD 2007.
[18] B. Feero, P.P. Pande, "Performance Evaluation for Three-Dimensional Networks-On-Chip", Proc. ISVLSI 2007.
[19] F. Li et al., "Design and Management of 3D Chip Multiprocessors Using Network-in-Memory", Proc. ISCA 2006, pp. 130-141.
[20] Y. Liu, et al., "Fine Grain 3D Integration for Microarchitecture Design Through Cube Packing Exploration", Proc. ICCD, 2007.
[21] D. Park et al. "MIRA: A Multi-layered On-Chip Interconnect Router Architecture", Proc. ISCA 2008, pp. 251-261.
[22] K. Puttaswamy, G. H. Loh, "Implementing caches in a 3D technology for high performance processors" Proc. ICCD 2005.
[23] I. Loi et al., "Supporting vertical links for 3D networks on chip: toward an automated design and analysis flow", Proc. NanoNet 2007.
[24] P. Koonath and B. Jalali, "Multilayer 3-d photonics in silicon", Opt. Express, 15(20):12686–12691, 2007.
[25] A. M. Pappu A. B. Apsel, "Analysis of intrachip electrical and optical fanout", Proc. Applied Optics, 44(30):6361–6372, Oct 2005.
[26] G. Tosik, et al., "Power dissipation in optical and metallic clock distribution networks in new VLSI technologies," Proc. IEE Feb 2004.
[27] I. O'Connor, "Optical solutions for system-level interconnect," SLIP, pp. 79-88, Feb 2004.
[28] D. Vantrease et al., "Corona: System Implications of Emerging Nanophotonic Technology," Proc. ISCA, pp. 153-164, 2008.
[29] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A.Choudhary, "Firefly: Illuminating future network-on-chip with nanophotonics," Proc. ISCA, pp. 429–440, 2009.
[30] A. Joshi et al., "Silicon-Photonic Clos Networks for Global On-Chip Communication", Proc. NOCS 2009.
[31] H. Gu et al., "A Low-Power Fat Tree-based Optical Network-on-Chip for Multiprocessor System-on-Chip", Proc. NOCS 2009.
[32] A. Shacham, K. Bergman, and L.P. Carloni, "The Case for Low- Power Photonic Networks on Chip," Proc. DAC, pp. 132-135, 2007.
[33] I. Artundo, et al, "Low-Power Reconfigurable Network Architecture for On-Chip Photonic Interconnects", Proc. HPI 2009.
[34] S. Bahirat, S. Pasricha, "Exploring Hybrid Photonic Networks-on-Chip for Emerging Chip Multiprocessors", IEEE/ACM CODES+ISSS Oct 2009
[35] I. Loi et al., "Synthesis of Low-Overhead Configurable Source Routing Tables for Network Interfaces", Proc. DATE 2009.
[36] M. A. A. Faruque, "Configurable Links for Runtime Adaptive On-chip Communication", Proc. DATE 2009.
[37] M. Li, et al., "DyXY - A Proximity Congestion-Aware Deadlock-Free Dynamic Routing Method for Network on Chip", Proc. DAC 2006.
[38] B. R. Koch, A. W. Fang, O. Cohen, and J. E. Bowers, "Mode-locked silicon evanescent lasers," Opt. Express 15(18), Sep 2007.
[39] A. Gupta, et al., "High-Speed Optoelectronics Receivers in SiGe," Proc. Intl. Conference on VLSI Design, pp. 957-960, Jan 2004.
[40] S.C. Woo et al."The SPLASH-2 programs: Characterization and methodological considerations", ISCAS, 1995.
[41] J. Rabaey et al., "Digital Integrated Circuits", Prentice Hall, 2002.
[42] S. Pasricha, and N. Dutt. "On-Chip Communication Architectures", Morgan Kauffman, Apr 2008
[43] G. Chen, et al., "Predictions of CMOS Compatible On-Chip Optical Interconnect," Proc. of SLIP, pp. 13-20, 2005.
[44] I.-W. Hsieh, et al., "Ultrafast-Pulse Self-Phase Modulation and Third-Order Dispersion in Si Photonic Wire-Waveguides," Opt. Exp., 2006.
[45] S. Pasricha, "Exploring Serial Vertical Interconnects for 3D ICs", IEEE/ACM Design Automation Conference (DAC), Jul 2009
[46] A. Kahng, et al., "ORION 2.0: A Fast and Accurate NoC Power and Area Model for Early-Stage Design Space Exploration" DATE, 2009.
[47] H. Al-Awwami, M. S. Obaidat, and M. Al-Mulhem, "ZOMA: A Preemptive Deadlock Recovery Mechanism for Fully Adaptive Routing in Wormhole Networks", Proc. ICCNMC, pp. 519-525, 2001.