Galaxy: A High-Performance Energy-Efficient Multi-Chip Architecture Using Photonic Interconnects

Yigit Demir†, Yan Pan*, Seukwoo Song‡, Nikos Hardavellas†, John Kim‡, and Gokhan Memik†

†Northwestern University
Dept. of Electrical Eng. and Computer Science
Evanston, IL, USA
yigit@u.northwestern.edu
{nikos, g-memik}@northwestern.edu

*Globalfoundries Inc.
Malta, NY, USA
panyan@gmail.com

‡KAIST
Dept. of Computer Science
Daejeon, Korea
jjk12@kaist.edu
sukwoo24@gmail.com

ABSTRACT

The scalability trends of modern semiconductor technology lead to increasingly dense multicore chips. Unfortunately, physical limitations in area, power, off-chip bandwidth, and yield constrain single-chip designs to a relatively small number of cores, beyond which scaling becomes impractical. Multi-chip designs overcome these constraints, and can reach scales impossible to realize with conventional single-chip architectures. However, to deliver commensurate performance, multi-chip architectures require a cross-chip interconnect with bandwidth, latency, and energy consumption well beyond the reach of electrical signaling.

We propose Galaxy, an architecture that enables the construction of a many-core "virtual chip" by connecting multiple smaller chiplets through optical fibers. The low optical loss of fibers allows the flexible placement of chiplets, and offers simpler packaging, power, and heat requirements. At the same time, the low latency and high bandwidth density of optical signaling maintain the tight coupling of cores, allowing the virtual chip to match the performance of a single chip that is not subject to area, power, and bandwidth limitations. Our results indicate that Galaxy attains speedup of 2.2x over the best single-chip alternatives with electrical or photonic interconnects (3.4x maximum), and 2.6x smaller energy-delay product (6.8x maximum). We show that Galaxy scales to 4K cores and attains 2.5x speedup at 6x lower laser power compared to a Macrochip with silicon waveguides.

Categories and Subject Descriptions

C.1.2 [**Computer Systems Organization**]: Multiprocessors—*Interconnection architectures*; B.4.3 [**Hardware**]: Interconnections—*Topology;* C.1.4 [**Computer Systems Organization**]: Parallel architectures

Keywords

Interconnection Networks; Nanophotonics; Energy Efficiency

1. INTRODUCTION

The physical limitations in area, yield, off-chip bandwidth, and power, limit the scalability of single-chip designs. Area and yield considerations push for small die sizes, and the latest ITRS [8] models reflect the competitive requirements for affordability by targeting flat chip-size trends for both cost-performance and high-performance processors (140-260 mm^2). At the same time, while transistor counts grow exponentially, voltage scaling has slowed. This has lead to a dramatic increase in power density with decreasing feature size [11], creating chips that require a power budget beyond what is practical today to operate and leading to "dark silicon" [7,9,17]. In addition, the limited pin count and low efficiency of off-chip communication severely limit the off-chip bandwidth [23], and hamper the scalability and performance of future multicores, even for highly-parallel workloads [9].

Physical constraints limit single chip designs to either a relatively small number of cores, beyond which scaling becomes impractical, or to designs that trade single-core performance for high aggregate instruction throughput, which can only be achieved if all cores are simultaneously employed by the executing workload. For example, a single core in Intel i7-3960X has a peak theoretical performance of 187 *GFLOPS*, but only 6 such cores fit in the chip's area and power budget. In contrast, Intel Phi 5110P features 60 cores, but at only 17 *GFLOPS* per core, and NVIDIA GTX-680 features 1536 CUDA cores but at a paltry 2 *GFLOPS* each.

Aggregating together several discrete smaller dies instead of having a large one (*disintegration*) overcomes the area and yield limitations, as only few dies need to be replaced if they are faulty [3,5]. The total silicon area of the aggregate chip can even scale beyond reticle size limits, allowing the aggregate "virtual" chip to reach scales impossible to realize with a monolithic design (*macrochip integration*). At the same time, a monolithic design forces the use of a technology that is only the best average for all circuit applications that share the die (e.g., cores may strive for a small and fast 16nm process, while an analog component may be more economic to stay at an old 90nm technology node). Disintegration allows each application to optimize its technology independently, reaching a better global optimum design point. However, disintegration and macrochip integration may come at the cost of increased power density, if the dies reside within the same package, or the high energy and latency cost of communication across discrete chips. High power density would force the chips to run at a lower power budget to stay within the thermal envelope, hampering scalability and degrading performance. High-latency high-energy cross-chip communication would render large-scale macrochips impractical. An ideal design would **(a)** mitigate the area and yield constraints, **(b)** achieve a low power profile, and **(c)** provide high-bandwidth, low-latency and low-energy communication across the discrete dies, while **(d)** maintain high scalability. Known alternatives can break free of some physical limitations and satisfy some of these requirements, but not all.

ICS'14, June 10–13, 2014, Munich, Germany.

ACM 978-1-4503-2642-1/14/06…$15.00.
http://dx.doi.org/10.1145/2597652.2597664

3D-die stacking can ease the area and yield limitations by vertically connecting several smaller dies in a package with through-silicon-vias (TSVs). However, 3D-die stacking incurs significant challenges in power delivery and heat removal, and is best employed when the additional dies implement low-power applications (e.g., DRAM). Thus, 3D integration fails to provide low power density and high scalability (requirements b, d).

Silicon interposers (i.e., 2.5D integration) allow chips to connect laterally within the same package through "bridge" silicon chips, thus exploiting the high density of die-to-package and on-chip wires. However, interposers enable only small-size arrays of chips that can fit in a single package. Scalability is further limited by the low speed of on-chip wires in distances over 10 *mm* [13,14], and the cooling limitations of a single package. Thus, while silicon interposers run cooler than a 3D-integrated design, they still fail to provide low power density and high scalability (requirements b, d).

Electrical links suffer from severely constrained bandwidth, due to limitations in the density of chip I/O and package routes, which dramatically constrain the number of links that can be routed across discrete chips. A 580 mm^2 die can have 25600 pins to the package substrate at a pitch of 150 μm, but the substrate-to-board pitch is 0.8 *mm*, which allows only 3844 pins to the board from a 5 *cm* x 5 *cm* package [8]. This forces the use of over-clocked and high-power serial links across chips. Unfortunately, electrical links driven by a high-speed serializer/deserializer circuit (*SerDes*) [22] on an FR-4 board incur significant energy consumption or long delays (typically 20 *pJ/bit*, and at best 2.5 *pJ/bit* and 2.5 *ns* latency over 4 inches of electrical strip [22]) as the designers have to trade energy for performance. Thus, SerDes links enable only a small array of chips, and fail to satisfy requirements c and d. To avoid confusion with on-chip wires, in this paper we use "*SerDes*" to refer to conventional electrical links across chips.

With the introduction of nanophotonics, systems can break free of all these limitations. The low latency and high bandwidth density of optical signaling can facilitate efficient off-chip communication and bring physically distant chips effectively close together. This makes it possible to build a physically-large but logically-dense many-core "virtual chip" by optically connecting several chiplets together [3,13,19], each with its own separate package and cooling.

To integrate chiplets into a larger system, NSiP [5] uses silicon-nitride waveguides across chiplets within a package, and the Oracle Macrochip [13] uses silicon waveguides etched on a wafer. While these proposals mitigate the area, yield, and memory bandwidth limitations of conventional designs, they do not address the power constraints. Thereby, designs utilizing waveguides are confined to a small physical space (e.g., a wafer [13] or a package [5]). This increases the thermal density to the point where liquid cooling is required to avoid thermal runaways [13,14], or confines the aggregate "virtual chip" to power limitations not much different from a monolithic design [5]. In addition, the optical loss of silicon waveguides (typically 0.05-0.3 *dB/cm* [14,4]) makes routing long cross-chiplet optical channels impractical. Macrochip integration may require links over 45 *cm* [14], for which waveguides impose a 5 *ns* latency and 2.25 *dB* loss. Ultra-low-loss waveguides can be manufactured, but their high area occupancy may result in exceedingly narrow chiplet-to-chiplet links (e.g., 2-bit links for an 8x8 chiplet array [13,14]) which in turn impose significant serialization that degrades performance. Thus, to design a large "virtual chip" using waveguides, one has to suffer high optical loss which multiplies the power requirements, or narrow paths which impose serialization, hurt performance, and in turn increase overall energy consumption.

In contrast, Galaxy is designed to push back the power limits, in addition to overcoming the area, yield, and bandwidth constraints, while allowing highly-scalable designs. Optical fibers have tremendously low optical loss (0.2 *dB/Km*), so very long channels can be drawn at very low power. Galaxy capitalizes on this and uses fibers for cross-chiplet communication. Its design also guarantees that each optical path employs only a small number of couplers, keeping the total optical loss and the corresponding laser power low. These two design choices allow spreading discrete chiplets far apart in space to minimize heat transfer and lower the power density of the virtual chip, which in turn enables each chiplet to operate at higher frequency than power-limited designs. At the same time, the propagation speed of light in fibers (0.676 *c)* is considerably higher than in silicon waveguides (0.286 *c*), allowing for low-latency long-distance communication. Compared to SerDes lines, fibers transmit at 33x lower energy per bit [2]. Thus, fibers provide high bandwidth at low power, and enable highly-scalable designs.

Previous research [13] dismissed the use of optical fibers for cross-chiplet communication under the assumption that chips connect to fibers at a relatively large 250 μm core pitch, not the 20 μm pitch of optical proximity couplers that silicon waveguides use. Hence, the chip-to-chip bandwidth over fibers would not improve much over area solder balls connected to package routes. Galaxy overcomes this limitation by exploiting a recently-demonstrated technology that couples an array of fibers into an array of waveguides at 20 μm pitch at the edge of the chip [15]. Our results indicate that fibers provide sufficient bandwidth for communication to chiplets and to memory, allowing for much wider data paths than low-loss but slow silicon waveguides, and boost both the performance and the energy efficiency of the multi-chip system by several times.

In summary, optical fibers are faster, impose lower optical loss, and require lower energy than available alternatives for chiplet communication. They are also flexible and allow for arbitrary placement of chiplets (e.g., across boards within a rack) without additional coupling. Thus, fibers are especially suitable for long, inter-chiplet optical channels, as they are easy to route, and can go off the plane or off the board. Galaxy utilizes optical fibers for cross-chiplet communication and offers simple packaging, power, and heat requirements, yet provides the performance advantages of a tightly-coupled system. While prior works have touched upon some of these issues in the context of multi-chip architectures [2, 3, 5, 13, 14, 19], to the best of our knowledge, this is the first work that quantifies the impact of disintegration and multi-chip integration on power constraints, and analyzes the performance, power, energy, and thermal behavior of multi-chip design alternatives.

It is important to note that Galaxy is just one design that supports processor disintegration and macrochip integration. Other topologies and designs are possible. Our goal is not to perform a full design-space sweep and advocate Galaxy as the optimal solution. Rather, we aim to demonstrate that processor disintegration can match the performance of designs that are not limited by power and off-chip bandwidth, effectively breaking free from the limitations of today's monolithic chips, and at the same time support large-scale macrochip integration. Specifically, our contributions are:

1. We quantify the impact of power and bandwidth constraints in monolithic single-chip designs, and the limitations of electrical links and SOI waveguides when used for chip communication.

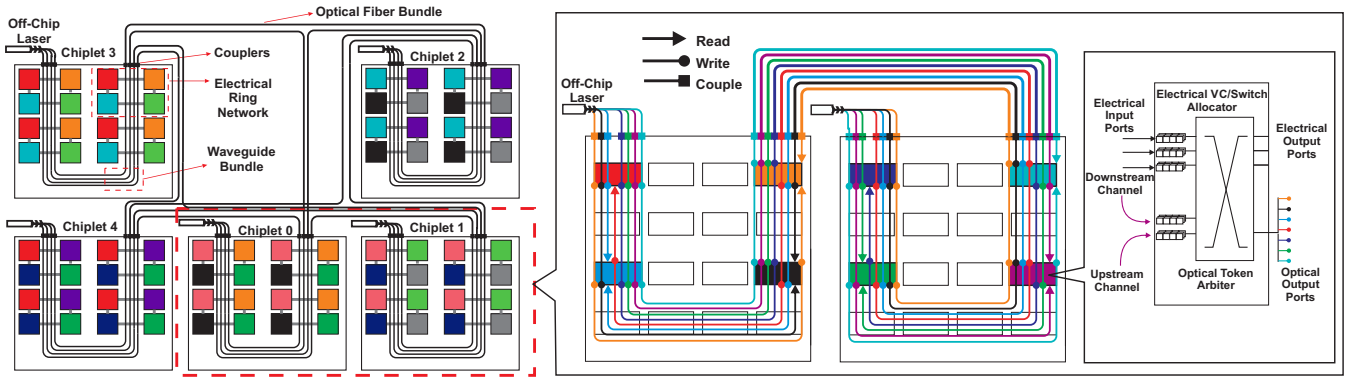

FIGURE 1. (a) Galaxy layout of an example 5-chiplet design, (b) MWSR optical crossbar, and (c) router architecture.

2. We propose Galaxy, an architecture that allows both processor disintegration and macrochip integration. Galaxy builds a many-core "virtual chip" by connecting multiple smaller chiplets through optical fibers.
3. We evaluate the performance, power, energy, and thermal profile of Galaxy, and compare it against single-chip designs (*processor disintegration*) and multi-chip designs (*macrochip integration*). Galaxy is up to 3.4x faster (1.8-2.2x on average) over single-chip alternatives with electrical, photonic, or hybrid interconnects, achieves up to 6.8x smaller energy-delay product (2.6x on average), and scales to 4K cores while being 2.5x faster at 6x lower laser power than a waveguide-based design.

2. THE GALAXY ARCHITECTURE

Galaxy builds a physically-large but logically-dense many-core "virtual chip" by optically connecting many discrete chiplets together. Galaxy aims for high performance while providing high energy efficiency and high scalability. Galaxy builds a point-to-point chip-to-chip network, which outperforms other switched networks in terms of performance and energy efficiency [13]. The chip-to-chip photonic interconnect extends across chiplets by coupling light to SOI waveguides from optical fibers at the edge of the chip [15]. Within a chiplet, Galaxy utilizes electrical signaling for nearest-neighbor communication, and silicon waveguides for long-distance communication. Long-distance on-chip communication has been shown to be more energy efficient with photonics than electrical signaling [21]. Furthermore, we will show that a seamless extension of on-chip optical signaling to chip-to-chip links allows Galaxy to connect an array of distributed chips at a performance comparable to on-chip interconnects [21,28] (Figure 5).

2.1 Network Topology

Galaxy employs a hybrid electrical/photonic interconnect. It extends Firefly [21] to support cross-chiplet communication at low power by minimizing coupler crossings and the number of sharers of each optical path. Figure 1(a) depicts an example 5-chiplet Galaxy design. The colored squares within each chiplet represent routers. The routers within a chiplet are divided into local clusters. Each cluster contains exactly one router per remote chiplet. In our example, there are 4 clusters per chiplet, with 4 routers per cluster. A local cluster in Chiplet 3 consists of neighboring red, orange, blue, and green routers (red outline in Chiplet 3, Figure 1(a)). Each cluster supports a number of cores based on a concentration factor. The cores and routers in a cluster are electrically connected. In our example, we use concentration 1 and an electrical ring within the cluster (other topologies are possible). A source-destination pair within the same cluster uses only electrical links.

Clusters communicate with each other through optical crossbars. Every optical crossbar is represented by coloring routers with the same color. For example, the pink routers in Chiplet 0 and the pink routers in Chiplet 1 belong to the same optical crossbar. Each optical crossbar extends across only two chiplets. This minimizes coupler crossings and optical loss: every optical path is short, and has at most 3 couplers (including the laser coupling). This way, Galaxy forms a fully connected point-to-point network between chiplets. Also, every crossbar extends across all clusters of the two chiplets it connects. In Figure 1(a), the crossbar between Chiplet 0 and Chiplet 1 consists of the pink routers in Chiplets 0 and 1, the U-shaped waveguides that connect these routers within each chiplet, and the fibers that connect the two chiplets. Figure 1(b) shows a close-up of that crossbar, where the pink routers have been re-colored to assist the detailed explanation of the crossbar later in the section.,

Routing a packet from Chiplet 0 to Chiplet 1 is carried by traversing the corresponding optical crossbar. This is done in 3 steps: (1) Route electrically within the source cluster in Chiplet 0 to a pink router; (2) Take the optical link and arrive at the pink router of the destination cluster in Chiplet 1; (3) Route electrically within the destination cluster to the destination core. Communication between any two clusters is performed similarly. Source-destination cluster pairs within the same chiplet use only the silicon waveguides in that chiplet. If the clusters are at different chiplets, the packet will traverse the waveguides within the source chiplet, the fiber connecting the two chiplets, and the waveguides in the destination chiplet. A packet that traverses an optical link will directly reach a router within the cluster of the destination core, and every packet traverses the optical link only once.

In general, if each chiplet has X clusters, each with Y routers, and a concentration of c, the proposed Galaxy architecture can connect $(Y+1)$ chiplets, using radix-$(2X)$ optical crossbars, supporting a total of $c*Y*X*(Y+1)$ cores. The example in Figure 1 is a case with $X=Y=4$, $c=1$, for a total of 80 cores. It is important to note that it is easy to extend Galaxy to support an arbitrary number of chiplets by having optical routers belong to multiple optical crossbars. However, for ease of explanation, we refrain from this design choice.

Firefly [21] uses Single Writer Multiple Reader (SWMR) optical crossbars, which use global broadcast channels to reserve the data channel, and requires an optical credit stream to control buffer space, thereby increasing power consumption. Galaxy adopts a modified Firefly topology with Multiple Writer Single Reader

(MWSR) optical crossbars which only require a token stream to manage arbitration and buffer space. In MWSR crossbars, each router "listens" on a dedicated channel and sends flits on the listening channels of all the other routers in the crossbar. Figure 1(b) illustrates the MWSR crossbar that extends over chiplets 0 and 1, with 8 senders and 8 receivers. The participating routers were shown in pink color in Figure 1(a), but they are shown with a distinct color here to ease explanation. Every router receives data from its own channel, which is shown with the same color as the receiver router, and writes 7 other channels which are the listening channels of the other routers in the crossbar. Galaxy adopts a 1-pass optical token stream with FairQuota [20] to guarantee that only a single router transmits on a channel at any moment, avoid starvation and packet loss due to buffer overflow, and provide QoS.

Because the optical links are traversed at most once, two Virtual Channels (VCs) are sufficient for the optical channels. The buffers of each optical VC are arbitrated using a separate optical VC token stream. To keep the balance of tokens, the tokens perform a double traversal. The receiver router of a channel first sends the VC tokens in the direction opposite to the data channel (back-traversal), all the way to the origin of the laser injection point, skipping all the senders on the way. Then, the VC token goes through O/E and E/O conversion, and is re-modulated onto a VC token stream in the same direction as the data channel (forward-traversal).

Figure 1(c) shows a hybrid electrical/optical router in Galaxy. Routers store the flits received from the electrical or optical networks in electrical buffers, after an optical to electrical (O/E) conversion if needed. Two electrical input and output ports route packets on the electrical local cluster ring. The third electrical input and output port is used for data injection. Each router has a pair of dedicated optical receiving channels, the upstream and downstream channels. The dark blue and green routers in Figure 1(b) send messages to the purple router through its upstream channel, while the rest send messages to the purple router through its downstream channel. Thus, 2 extra ports are added on the input side of the router to receive packets from the dedicated optical receiving channels from both directions. On the output side, 7 additional output ports switch outgoing packets to different optical channels.

2.2 Inter-Chiplet Connection

Galaxy targets large-scale macrochip integration (e.g., 60+ chiplets) which require fast and energy-efficient long-distance links. Thus, Galaxy connects chiplets via optical fibers, rather than SerDes links [22], or silicon waveguides (0.05 *dB/cm* [14]). SerDes provide at best 2.5 *pJ/bit* and 2.5 *ns* latency over 4 inches of electrical strip [22], thus fibers offer lower latency, and two orders of magnitude lower energy/bit and energy-delay-product across the entire range of possible chiplet-to-chiplet distances (Figure 2). Similarly, fibers are almost 2x faster than SOI waveguides, and achieve between 2-10x lower energy/bit and energy-delay-product

TABLE 1. Nanophotonic Parameters

	per Unit	Total
Splitters	0.2 *dB*	0.2 *dB*
Waveguide Loss	0.3 *dB/cm*	1.5 *dB*
Fiber Loss	0.2 *dB/Km*	~0 *dB*
Nonlinearity	1 *dB*	1 *dB*
Coupler Loss	3.8 *dB*	7.6 *dB*
Modulator Insertion	0.5 *dB*	0.5 *dB*
Ring Through	0.01 *dB*	1.28 *dB*
Filter Drop	1.5 *dB*	1.5 *dB*
Photodetector	0.1 *dB*	0.1 *dB*
Total Loss		**13.68 *dB***
Detector Sensitivity	**-20 *dBm***	
Modulation/Demodulation	**150 *fJ/bit***	
Laser Power per Wavelength		**0.233 *mW***
Total Laser Power		**1.195 *W***

(Figure 2), mainly due to the high relative optical loss and refractive index of typical silicon waveguides. Fibers are especially suitable for long, inter-chiplet channels, allowing Galaxy to have a thermal-aware design while maintaining high performance and energy efficiency. Figure 2 corroborates prior research [14].

Fibers connect to chiplets through a coupler that tapers an array of fibers at 250 μm pitch down to 20 μm pitch channels, and couples them into an array of SOI waveguides at the edge of the chip [15]. The measured coupling loss is 3.8 *dB,* including tapering the channels, the refraction index change from fibers to the waveguides, and misalignment [15]. Misalignment within 0.7 μm, 0.4 μm, and 0.7 μm in the lateral, vertical, and optical axes produces losses under 1 *dB* [15]. The performance of the tapered coupler is comparable to that of an optical proximity coupler (3.5 *dB* coupler loss, plus 0.5 *dB* per 1 μm misalignment in the y-axis, plus less than 1 *dB* loss due to misalignment of 2.5 μm in the x- and z-axis [31]).

2.3 Nanophotonic Parameters and Power

On-chip lasers dissipate a lot of power and heat the chip, thus Galaxy adopts off-chip WDM-compatible lasers. The laser is brought on chip via fibers connected to tapered couplers [15], and a splitter distributes it to on-chip waveguides [4]. Tapered couplers also connect the on-chip waveguides to the off-chip optical fibers. Galaxy uses the modulators, demodulators, drop filters, splitters, and photodetectors introduced in [1]. Optical links run at 10 *GHz* and are implemented similar to [1]. Table 1 details the optical parameters.

The example configuration of the 5-chiplet Galaxy we evaluate in Section 4.2 consists of 10 radix-8 MWSR crossbars that transfer 64-bit flits. We assume a modest 16-way DWDM, thus Galaxy uses

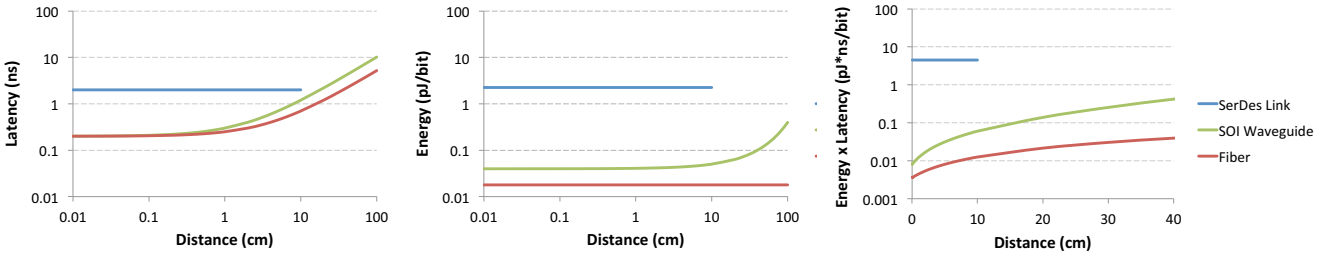

FIGURE 2. Latency, Energy / bit, and Energy x Delay product for SerDes links, SOI waveguides, and fibers.

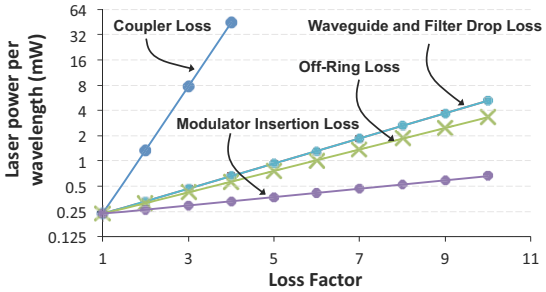

FIGURE 3. Laser power sensitivity to optical parameters.

a total of 320 fibers (128 fibers attached to each chiplet) and 40960 ring resonators (8192 per chiplet covering ~0.9 mm^2). Because every optical channel requires a 1-token-pass arbitration mechanism, a total of 20 additional fibers and 3840 rings are used for arbitration. Another 80 rings and 10 fibers are used for forward clock distribution [14].

To calculate the total ring heating power we extend the method by Nitta *et al.* [18] by incorporating the heat generated by the cores. The cores heat up the photonic layer, and the ring heaters provide the remaining heat necessary to bring the photonic layer within the ring tuning range. As current injection may cause a thermal runaway [18], we only consider trimming by heating. Section 3.1 details the model. We also include the trimming power required for process variations [12]. While Galaxy may benefit from trimming power saving methods [18], they are out of the scope of this paper.

Figure 3 demonstrates the sensitivity of Galaxy's laser power to a change in the loss of each nanophotonic parameter. The laser power is most sensitive to the coupler loss, but relatively insensitive to the other parameters, indicating that our results will likely hold under a wide range of nanophotonic device technologies.

When evaluating laser power consumption, existing literature typically omits inefficiencies in the generation and delivery of the laser [1,13,14,21,28]. By analogy, and to ease comparisons with prior work, we did not include the generation and delivery cost in the laser power calculations in Table 1 and the remainder of this paper. For completeness, however, we report here the laser power including all these overheads. The additional coupling loss increases the laser power to 2.9*W*. Given 10% efficiency for the WDM-compatible laser [32], the wall-socket laser power is 29*W*.

3. EXPERIMENTAL METHODOLOGY

We evaluate the performance of Galaxy for processor disintegration by modeling an example 5-chiplet 80-core Galaxy on a full-system cycle-accurate simulator based on Flexus 4.0 [10,29] integrated with Booksim 2.0 [6] and DRAMSim 2.0 [24]. Table 2 details the architectural modeling parameters. We target a 16 *nm* technology, and have updated our tool chain accordingly based on ITRS projections [8]. We follow the SimFlex sampling methodology [29] with 95% confidence intervals. The simulated system executes a selection of SPLASH and scientific workloads.

We compare Galaxy against three single-chip multicores, all of which implement the architecture described in Table 2. The first multicore uses an all-electrical 2D-Concentrated Mesh on-chip interconnect with express links [6] and concentration of 4 (**CMesh-Exp**). Concentrated mesh is often chosen for on-chip networks as it maps well to a 2D-VLSI planar layout with low complexity. We evaluated a regular 2D-Mesh and a 2D-Concentrated Mesh without express links, and found that CMeshExp outperforms the other

TABLE 2. Architectural Parameters.

Multicore Size	80-cores, 580 mm^2
Processing Cores	ULTRASPARC III ISA, max 5 *GHz*, OoO, 4-wide dispatch/retirement, 96-entry ROB
L1 Cache	split I/D, 64 *KB* 2-way, 2-cycle load-to-use, 2 ports, 64-byte blocks, 32 MSHRs, 16-entry victim cache
L2 Cache	shared, 512 *KB* per core, 16-way, 64-byte blocks, 14-cycle hit, 32 MSHRs, 16-entry victim cache
Memory Controllers	One per 4 cores, or 4 MCs per chip. 1 channel/MC Round-robin page interleaving
Main Memory	DDR3, 80 *GB*, 8 *KB* pages, 20 *ns* access latency Interfaces: (a) Conventional pins, (b) Optically-connected memory (OCM) [1], (c) 3D-stacked [13]
Networks	CMesh, Corona, Firefly, Galaxy, Oracle Macrochip

designs on all metrics (performance, power, and energy). Thus, we only show results for CMeshExp. We model routers with 8 input and output ports and a 3-cycle routing delay. Routers are connected through 166-bit bi-directional links with a 1-cycle link delay.

The second multicore uses an all-optical MWSR crossbar (**Corona** [28]), implemented with 80 MWSR optical busses (256-bit data channels). We model global switch arbitration using an optical token ring. A token for each node, which represents the right to modulate on the node's wavelength, continuously passes around all nodes on a dedicated arbitration waveguide. A node grabs and absorbs a token to transmit a packet, and then releases the token to allow other nodes to obtain it. We estimate 16 *cm* long waveguides for the Corona chip, resulting in 8 cycles token round-trip time.

The third multicore implements a hybrid interconnect where clusters of electrically-connected cores are connected through an SWMR optical crossbar entirely on chip (**Firefly** [21], 256-bit flits, 8 *cm* waveguide, and 1-cycle reservation and electrical-link delay).

We model **Galaxy** with 1-cycle latency for processing an optical token request [20]. Each Galaxy router can initiate a maximum of 8 token requests per cycle, but can utilize at most 2 acquired tokens [20]. Galaxy uses 1-pass token stream arbitration for combined VC and channel arbitration. We estimate that the round-trip time of a token is also 8 cycles (8 *cm* SOI waveguide plus 16 *cm* optical fiber travel time). The input buffers are implemented as a DAMQ [27], with packets queued separately based on their destination. A data packet contains 512 bits, divided into eight 64-bit flits.

3.1 Power and Temperature Modeling

All systems we model employ Dynamic Voltage and Frequency Scaling (DVFS) to lower the voltage and frequency of a chip or chiplet when it reaches the limits of safe operational temperature (without loss of generality, we assume 90^oC). Figure 4 shows the flow diagram of our simulation tool chain. We collect runtime statistics from full-system simulations, and use them to calculate the power consumption of compute cores, caches, and memory controllers using McPAT [16], and the power consumption of the electrical and optical networks using DSENT [26] and the analytical model by Joshi *et al.* [12] respectively. Based on these estimates, we calculate the temperature of the chip and chiplet assemblies using HotSpot 5.0 [25] and FloTherm [30], a computational fluid dynamics tool that models the heat transfer between chiplets through air flow and convection. The estimated temperature is then

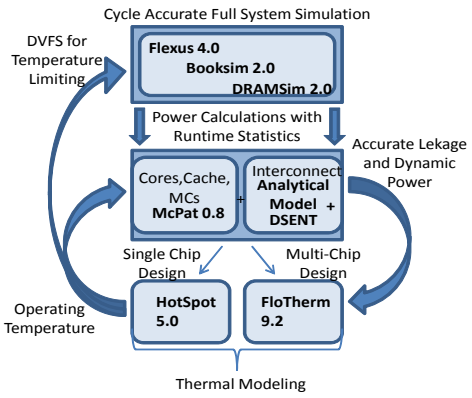

FIGURE 4. Simulation flow chart.

used to refine the leakage power estimate, and we iteratively calculate the power and temperature profiles until the system reaches a stable state. We use the stable-state power and temperature estimates to adjust DVFS, and repeat the process until we identify a DVFS setting for which the chip stays just below 90^oC, or operates at the maximum 5 *GHz*.

To calculate the total ring heating power for Galaxy, Corona, and Firefly, we extend the method by Nitta *et al.* [18] by additionally accounting for the heating of the photonic die by the operation of the cores. We model the thermal characteristics of a 3D-stacked architecture where the photonic die sits underneath the logic die using the 3D-chip extension of HotSpot [25]. For each target architecture (Corona, Firefly, and Galaxy) we measure the maximum temperature of the logic die during the execution of each one of the workloads. Then, we tune the micro-rings to the maximum of all the observed temperatures that the logic layer reaches across all benchmarks executing on the target architecture, plus a small margin. When a workload executes, we calculate the ring heating power required to maintain the entire photonic die at the micro-ring trimming temperature during the entire execution. We also include the trimming power required to overcome process variations [12].

3.2 Modeling Memory and Physical Constrains

To demonstrate the ability of disintegrated architectures to break free of power and bandwidth limitations, we evaluate Galaxy against all possible single-chip multicore combinations: power-constrained, off-chip bandwidth-constrained, fully constrained (i.e., both power- and bandwidth-constrained), and unconstrained.

We evaluate power-constrained multicores by employing DVFS to keep the chips within 90^oC. To evaluate multicores that are not subject to power constraints, we allow the chips to run at the maximum speed allowed by the design (5 *GHz*), by disregarding power and thermal limits. We evaluate bandwidth-constrained single-chip multicores by assuming a conventional DDR3 memory, and limit the total memory bandwidth by utilizing ITRS [8] pin projections for a 5 *cm* x 5 *cm* package, assuming 1/3 of the pins are used for power, 1/3 are used for I/O, and the remaining 1/3 are used for memory. The memory pins are distributed equally among four memory controllers (MCs). To evaluate designs that are not limited by memory bandwidth, we increase the number of pins well beyond ITRS projections and commensurately increase the number of MCs, until more pins or more MCs no longer increase performance. For our workloads, we reach this point when 5x more pins are distributed across 20 MCs. Fully constrained designs operate within the power, memory bandwidth, and thermal limits. Fully unconstrained designs operate beyond the power, thermal, and bandwidth limits and cannot realistically be built; however, they provide the highest performance that a particular architecture can achieve, limited only by the maximum speed allowed by the design (5 *GHz*). While we compare Galaxy to both constrained and unconstrained single-chip multicores, Galaxy is always modeled to conform to realistic power, bandwidth, and temperature limits.

TABLE 3. Galaxy scalability.

# of Cores	Multi-Chip Architecture		Bandwidth per Chip (TB/s)	Laser Power (W)	Serialization Overhead (cycles)	Link Latency (cycles)
320	Galaxy	Fibers	10	4.0	1	2
		Waveguides	5	4.9	2	10
		SerDes links	0.320	3.9	32	12
1088		Fibers	20	27.0	2	10
		Waveguides	5	26.0	8	20
		SerDes links	0.640	26.8	64	12
4160		Fibers	40	47.6	4	10
		Waveguides	10	44.9	16	20
		SerDes links	0.320	47.9	512	12
4096	Oracle MacroChip		0.630	~40.0	64	20

Emerging memory technologies (e.g., optically-connected memory (OCM) [1] or 3D-memory [13]) are not pin-limited, and can remove the memory bandwidth bottleneck for all multicore designs. Thus, we separately evaluate the performance of Galaxy against single-chip multicores with OCM and 3D-memory, where each multicore employs 20 MCs. We model a 10 *ns* access latency for OCM [1] and 2 *ns* for 3D-memory [13].

3.3 Modeling Large-Scale Multi-Chip Designs

Galaxy can scale up to 1088 cores with 17 chiplets (64 cores each with concentration 4), and 4160 cores with 65 chiplets. As the number of chiplets grows, the number of point-to-point links and their length increase. In order to keep the network power and component count within reasonable levels, off-chip bandwidth increases slower than the number of point-to-point links. As a result, the datapath width drops for large designs, forcing senders to send messages by sending many smaller flits serially (serialization overhead). Furthermore, messages take longer to travel the longer links. Galaxy does not suffer from high serialization or link delays and remains scalable, because it uses optical fibers, which provide high performance with energy efficiency (Section 2.2). We evaluate Galaxy for macrochip integration by comparing it against (a) **Galaxy with SOI** waveguides and optical proximity (OPC) couplers [31], (b) **Galaxy with SerDes** links, and (c) the **Oracle Macrochip** [13]. For fairness, we adjust the datapath width of Galaxy alternatives so they fit into similar power envelopes, and then we calculate the latency overhead. The Oracle Macrochip model closely follows [13,14,31]. Table 3 details the bandwidth, power

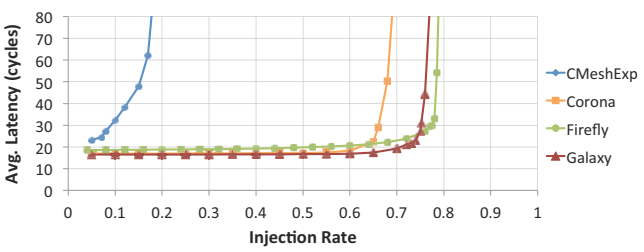

FIGURE 5. Load latency for uniform random traffic.

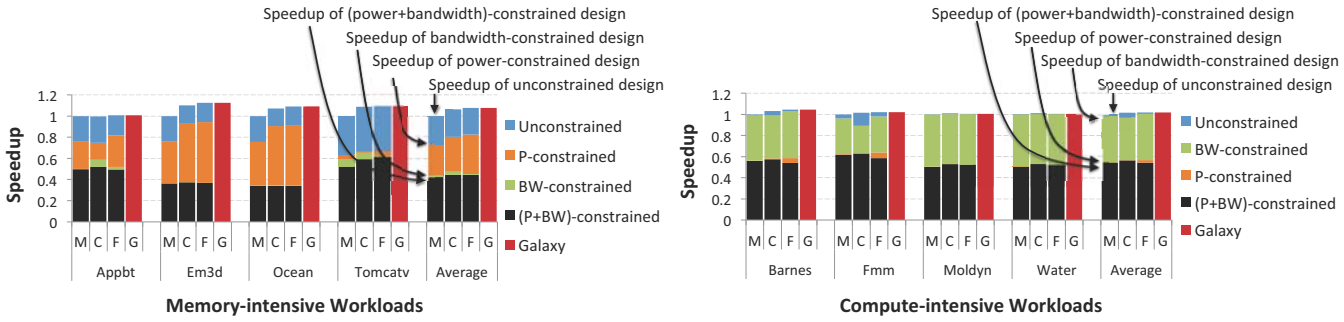

FIGURE 6. Speedup of constrained and unconstrained architectures: CMeshExp (M), Corona (C), Firefly (F), and Galaxy (G).

consumption, serialization delay, and link delay for all designs. To keep the simulations tractable, we estimate the performance of the scaled-out designs by imposing the latency overheads of each scaled-out system from Table 3 to an 80-core 5-chiplet model. The SerDes delay is optimistically kept constant for all sizes. As we impose the scaling overheads onto same-size designs in all cases (80 cores, 5 chiplets), the higher core count of Galaxy compared to the Oracle Macrochip does not affect the results.

4. EXPERIMENTAL RESULTS

4.1 Network Performance

Figure 5 analyzes the load-latency of CMeshExp, Corona, Firefly, and Galaxy. CMeshExp saturates quickly, which is indicative of its relatively low bandwidth. Corona saturates at a little less than 0.7 injection rate, while Firefly reaches an injection rate of almost 0.8 before saturating. Galaxy trails Firefly closely, and falls only slightly short in performance. This is expected because Galaxy is similar to a 2-level Firefly that creates a single datapath between two clusters, while packets in Firefly can take several alternate routes and utilize more of the available bandwidth. Nonetheless, the results indicate that Galaxy is a competitive interconnect.

4.2 Processor Disintegration

Figure 6 shows the speedup achieved by unconstrained single-chip designs (top of blue bar) with CMeshExp, Corona, and Firefly interconnects for memory-intensive and compute-intensive workloads. Submitting the multicores that run compute-intensive workloads (Figure 6 right) to realistic bandwidth constraints results in lower performance, but the loss is relatively small (top of green bar). Submitting them to power constraints, however, results in significant performance drop (top of orange bar). These multicores employ DVFS to stay below 90^oC, which slows down the compute-intensive workloads the most, as they have high core utilization which in turn dissipates more power. For example, Corona runs barnes at only 2.25 *GHz* from a nominal frequency of 5 *GHz*, and Firefly exhibits a similar slowdown. In comparison, Galaxy never exceeds 70^oC, and thus it can run at the full 5 *GHz* and outperform all single-chip alternatives by 1.8x on average. Multicores running memory-intensive workloads also show degraded performance when power-constrained (Figure 6 left, top of orange bar), indicating that power limitations are always an important factor. However, they incur the highest performance loss mainly when limited in off-chip bandwidth (top of green bar), while the slowdown due to DVFS is secondary. For example, CMeshExp runs em3d at 4.25 *GHz*, but Galaxy still demonstrates 3x speedup. Because of this dual slowdown, Galaxy achieves the maximum speedup over fully-constrained single-chip multicores (their performance is indicated by the top of the black bar) on memory-intensive workloads (2.3x on average, and up to 3.5x for ocean). More importantly, Galaxy manages to match or exceed the performance of designs that are entirely unconstrained. This demonstrates the ability of processor disintegration to break free of the power and bandwidth walls of conventional monolithic designs. All the designs we evaluate in the remainder of this paper are subject to power and off-chip bandwidth constraints, where the bandwidth limitations depend on the assumed memory technology.

Optically-connected memory (OCM) [1] overcomes the bandwidth limitations and decreases the memory latency. Corona with OCM outperforms Corona with conventional DDR3 by 3-4x on memory intensive workloads (Figure 7). Firefly and CMeshExp show similar trends. Galaxy, however, still outperforms all alternatives by 1.8x on average, as it runs at the full 5 *GHz* while DVFS limits the single-chip designs (e.g., Corona with OCM runs em3d at only 3.25 *GHz*). 3D-stacked memory has a similar effect on Galaxy, while Corona, Firefly, and CMeshExp do not get faster as they are still power limited. Overall, Galaxy outperforms alternative designs by up to 2.95x (2x on average). We conclude that Galaxy can leverage the emerging memory technologies to the fullest, while single-chip multicores are limited by the single-chip power envelope and fail to utilize fully the new memory technologies.

Figure 8 shows the breakdown of the normalized energy-delay product (EDP) and the average energy per instruction of CMeshExp, Corona, Firefly, and Galaxy with conventional memory. The

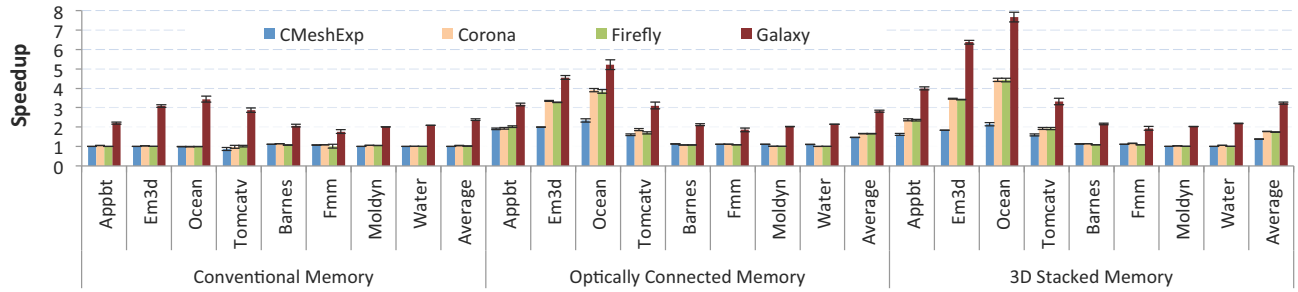

FIGURE 7. Speedup of power-constrained designs with various memory technologies (normalized to CMeshExp with DDR3).

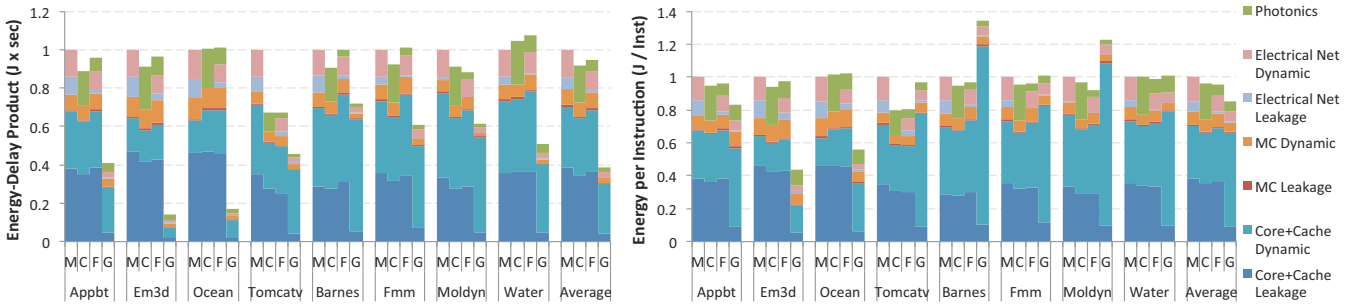

FIGURE 8. (a) Energy x Delay, and (b) Average energy / instruction for CMeshExp (M), Corona (C), Firefly (F), and Galaxy (G).

dynamic energy consumption of cores and caches for Galaxy is higher as it achieves 2.3x speedup on average over single-chip designs. This effect is more pronounced for compute-intensive workloads (barnes, moldyn). However, the chiplets in Galaxy run at only 70^oC and dissipate $55W$ each, compared to 90^oC and $130W$ for CMeshExp-, Corona-, and Firefly-based chips. As a result, Galaxy lowers leakage to just over 10% of energy, while single-chip designs waste 36-40% of their energy on leakage. Overall, single-chip designs consume 1.12-1.2x more energy per instruction than Galaxy (Figure 8(b)). Galaxy reaches its highest energy efficiency increase on memory-bound workloads (2-2.3x), as it achieves over 3x speedup and the chiplets dissipate less power waiting for memory. Galaxy attains up to 6.8x lower EDP than single-chip multicores (2.8x on average; Figure 8(a)).

Because Galaxy chiplets run cooler when running memory intensive workloads, the energy consumption of the photonics (including laser power, modulation/demodulation, and ring heating) is higher, as the ring heaters dissipate more power to keep the photonics layer at the trimming temperature. The ring heaters work less with compute intensive workloads, because cores dissipate more power and heat the photonic die.

4.3 Macrochip Integration and Scalability

Galaxy can scale up to 1088 cores with 17 chiplets, and 4160 cores with 65 chiplets (Section 3.3). We evaluate the scalability of Galaxy by comparing it against (a) Galaxy with SOI waveguides and OPC couplers [31], (b) Galaxy with SerDes links, and (c) the Oracle Macrochip [13] (Section 3.3). Table 3 details the power, bandwidth, and latency characteristics of the scaled out designs. Figure 9 compares the performance of these alternatives. The power-hungry SerDes links cannot provide enough bandwidth within the power envelope, resulting in high serialization delay that increasingly hurts performance as the system scales up. Similarly, SOI waveguides fall short because they require higher laser power than fibers, and at the same time light propagates 2.3x slower in waveguides. As a result, fibers increasingly outperform SOI waveguides as the system scales up. The performance gap is higher for memory-intensive workloads which stress the interconnect more. A 65-chiplet Galaxy with fibers outperforms Galaxy with SOI waveguides by up to 1.44x (1.24x on average), and Galaxy with SerDes by up to 9.53x (4.58x on average). The Oracle Macrochip [13,14] uses SOI waveguides and OPCs [31] to create point-to-point photonic links across chips. Galaxy outperforms the Oracle Macrochip by 2.5x on average (Figure 9) because the Macrochip implements a 2-bit-wide data channel with SOI waveguides, which impose high serialization and link delay.

We evaluate the sensitivity of laser power to the coupler loss for the Oracle Macrochip and Galaxy (Figure 10), because the coupler loss is the biggest contributor to the laser power consumption (Figure 3). We present laser power consumption of the Oracle Macrochip with measured coupler losses for passive-aligned and active-aligned OPCs [31], as well as under very aggressive OPC loss predictions of 1.2 *dB* [13,14]. For Galaxy, we present the laser power consumption under SION and SU8 tapered couplers using loss measurements of existing prototypes [15]. From one chiplet to another, laser has to pass through 3 couplers in Macrochip (vs. 2 for Galaxy), so the higher slope of the graph indicates higher sensitivity to coupler loss. The Macrochip with actively-aligned OPCs requires 6x more laser power than Galaxy. Even if the predicted OPC loss is achieved, Galaxy with existing couplers would still require less laser power.

4.4 Thermal Evaluation

To effectively push back the power wall while still employing conventional forced air cooling solutions and cheap packaging appropriate for high-volume markets, a disintegrated design requires the chiplets to be physically far enough from each other to minimize heat transfer. Our thermal modeling using computational fluid dynamics tools [30] and HotSpot [25] indicates that a Galaxy architecture with active heatsinks on each chiplet allows the chiplets to operate at 66.2^oC, sufficiently cool for most applications. In fact, even cheaper cooling solutions seem adequate. Figure 11(a)

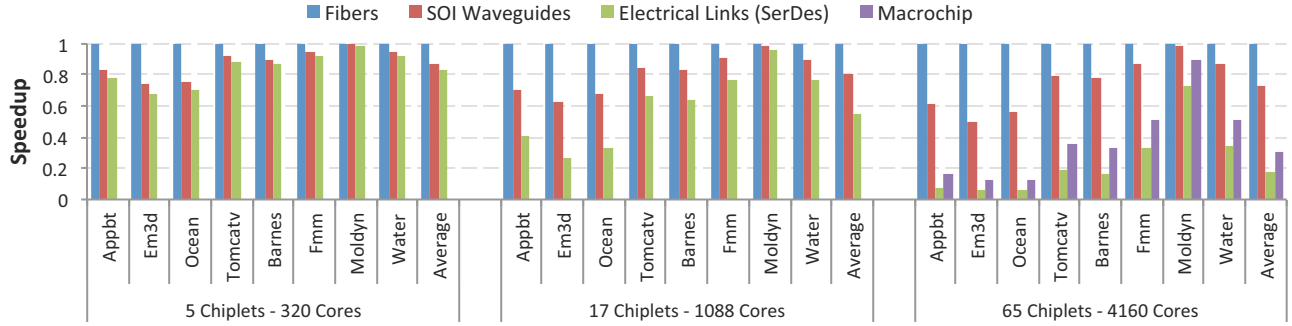

FIGURE 9. Comparison of Galaxy with different chiplet-to-chiplet interconnect technologies, and the Oracle Macrochip.

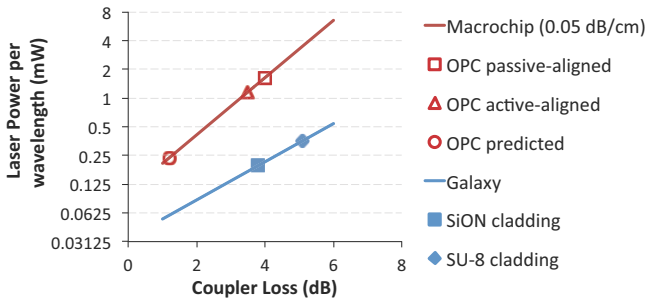

FIGURE 10. Laser power sensitivity to coupler loss.

shows a 5-chiplet Galaxy design which uses passive heatsinks spaced 8 *cm* apart, with a global fan blowing air horizontally in 45^oC ambient temperature in a box shell. The fanless (passive) heatsinks cool chiplets to 88.2^oC, and deliver low packaging and cooling costs, and increased lifetime. Thus, even very simple and cheap cooling solutions (fanless heatsinks and a global fan) suffice for an 80-core design disintegrated to 5 chiplets with Galaxy.

Optical fibers allow Galaxy to spread chiplets far apart for better cooling, while SOI waveguides and SerDes confine them to limited physical space (e.g. a wafer [13]). We compare the thermal behavior of a Macrochip-like dense design to an equal-size Galaxy by modeling a 9-chiplet design. Both designs use the same heatsinks and dissipate 50W per chiplet. Based on the Macrochip architecture [13,14], we estimate that the heatsinks will almost touch each other resulting in the layout shown at Figure 11(b). We observe that the sites that are further away from the fan reach 249^oC, and hence require a special cooling solution. A thermal-aware placement of 9 Galaxy chiplets on a 2D-plane (Figure 11(c)) achieves a maximum temperature of 110^oC, which is a full 139^oC lower than Macrochip. Furthermore, using optical fibers for cross-chiplet communication allows Galaxy to utilize multiple boards, in which case Galaxy can bring a 9-chiplet design down to a cool 87^oC (Figure 11(d)). This freedom of placement gives a significant advantage to Galaxy compared to silicon-waveguide-based designs, and allows it to spread the volume enough to cool even large-scale designs.

5. LIMITATIONS AND CHALLENGES

5.1 Coupler and Fiber Density Considerations

The use of fibers for chiplet-to-chiplet communication in Galaxy brings two new challenges: coupling the fibers on chip, and attaching enough fibers to achieve the highest performance or lowest EDP, depending on the optimization target. Galaxy requires enough length along the periphery of a chiplet to attach the fibers. Even the simulated 116 mm^2 chiplets provide over 43 *mm* in total length along the edge of a chip, allowing up to 172 fibers at a conservative 250 μm pitch. The disintegrated design we evaluated assumes 128 fibers per chiplet with 16 DWDM on 64-bit-wide datapaths. Figure 12 indicates that a higher fiber density provides only marginal performance benefits at significantly higher laser power (e.g., 512 fibers provide 3% more speedup at 4x more power), while fewer fibers reduce performance without significant power savings (e.g., 16 fibers are 15.5% slower and save only 1W). This indicates that the power and performance are the real limiting factors, not the fiber density provided by the coupler.

5.2 Yield and Financial Considerations

Galaxy relies on the manufacturing of a photonic die, the 3D integration of the photonic and the logic dies at each chiplet, the manufacturing of tapered couplers and fibers, and their assembly. The absence of yield and manufacturing data for nanophotonic systems does not allow us to make quantitative arguments, but we note that fibers have been manufactured at high volumes and have become relatively cheap. To assist in calculating the additional cost of nanophotonics, Section 2.3 provides component counts for the nanophotonic devices. Processor disintegration allows Galaxy to recover the cost overhead of nanophotonic devices by breaking a monolithic chip into multiple smaller chiplets, increasing yield and lowering non-recurring and marginal costs by a significant factor, as only the defective chiplets need to be replaced rather than an entire large chip ([5]). This is especially important for low and medium volume markets. Furthermore, by allowing for the arbitrary placement of chiplets, Galaxy offers the system architect the flexibility to utilize any point in the trade-off between compute density and cooling, from forced air to liquid cooling and beyond. Fibers allow the chiplets to spread in 3D-space and occupy multiple boards, balancing compute density with board power requirements and cooling, while still performing like a large monolithic tightly-coupled chip (Section 4.4). Even though photonic integration adds to the cost, Galaxy's higher yield, scalability, performance, and energy efficiency may overcome the extra cost and result in competitive and possibly even lower cost of ownership.

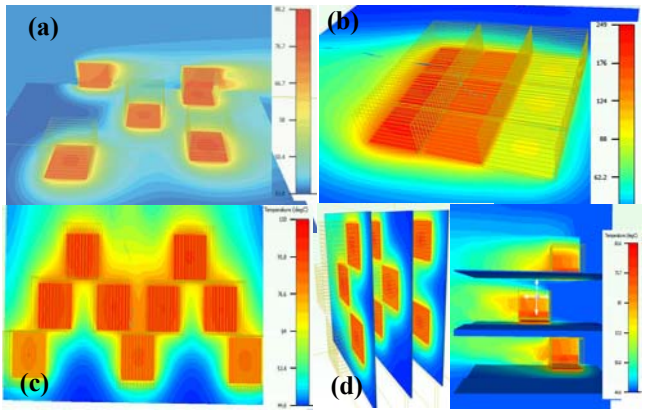

FIGURE 11. Thermal effects of chiplet placement.

6. RELATED WORK

Several on-chip interconnect networks exploiting optical signaling have been proposed [12, 21, 28]. Beamer *et al*. [3] proposed to achieve higher hardware parallelism while using smaller dies with high production yield. Batten *et al*. [1] proposes to connect a many-core processor to the DRAM memory using monolithic silicon. Koka *et al*. [13] discuss the design and implementation of a silicon-photonic network for a large multi-die "macrochip" system. In con-

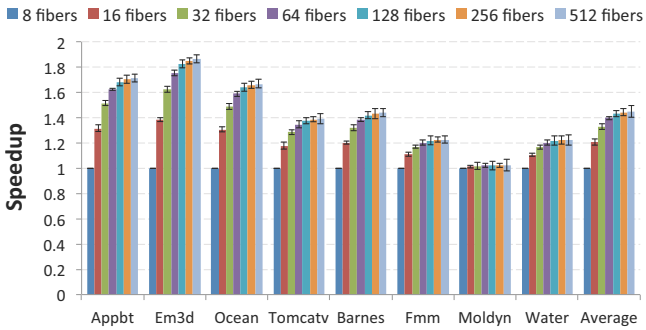

FIGURE 12. Sensitivity to fiber density per chiplet.

trast to these architectures, Galaxy leverages optical fibers to create a high-bandwidth, scalable, low-latency photonic interconnect that can support both processor disintegration and multi-chip integration, and at the same time enable cheap cooling solutions.

7. CONCLUSIONS

In this paper we propose Galaxy, a multi-chip architecture which builds a many-core "virtual chip" by connecting multiple smaller chiplets through optical fibers. Galaxy is designed to push back the power constraints, in addition to overcoming the area and bandwidth limitations, while providing high scalability. We demonstrate that Galaxy achieves 1.8-3.4x average speedup over competing single-chip designs, and achieves 2.6x lower energy-delay product (6.8x maximum). The careful design of optical paths in Galaxy minimizes coupler crossings, and allows it to scale beyond 4K cores, showing significant promise as the foundation of practical large-scale virtual chips. We show that a scaled-out 4K-core Galaxy attains significant speedup and energy efficiency advantages over similar designs such as the Oracle Macrochip, as it achieves at least 2.5x speedup with 6x more power-efficient optical links.

8. ACKNOWLEDGEMENTS

This work was partially supported by National Science Foundation awards CCF-1218768, CCF-0747201, and CCF-0916746, an ISEN booster award, and the June and Donald Brewer Chair in EECS at Northwestern University. John Kim was supported by the IT R&D program of MSIP/KEIT [10041313, UX-oriented Mobile SW Platform] and by the NRF grant funded by the Korea government (MSIP) (NRF-2013R1A2A2A01069132).

9. REFERENCES

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

[2] S. Beamer. *Designing Multisocket Systems with Silicon Photonics*. Ph.D. thesis, University of California at Berkeley, 2009.

[3] S. Beamer, K. Asanovic, C. Batten, A. Joshi, and V. Stojanovic. Designing multi-socket systems using silicon photonics. In *International Conference on Supercomputing (ICS)*, pages 521- 522, 2009.

[4] J. Cardenas, C. Poitras, J. Robinson, K. Preston, L. Chen, and M. Lipson. Low loss etchless silicon photonic waveguides. *Optics Express*, 17(6):4752- 4757, 2009.

[5] M. Cianchetti, N. Sherwood-Droz, and C. Batten. Implementing system-in-package with nanophotonic interconnect. *Workshop on the Interaction between Nanophotonic Devices and Systems*, 2010.

[6] W. J. Dally and B. Towles. *Principles and practices of interconnection networks*. Morgan Kaufmann Publishing Inc., 2004.

[7] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger. Dark silicon and the end of multicore scaling. In *38th Annual International Symposium on Computer Architecture*, 2011.

[8] European, Japan, Korean, Taiwan, and United States Semiconductor Industry Associations. The international technology roadmap for semiconductors (ITRS). http://www.itrs.net/, 2012 Edition.

[9] N. Hardavellas, M. Ferdman, B. Falsafi, and A. Ailamaki. Toward dark silicon in servers. *IEEE Micro*, 31(4):6- 15, July-August 2011.

[10] N. Hardavellas, S. Somogyi, T. F. Wenisch, R. E. Wunderlich, S. Chen, J. Kim, B. Falsafi, J. C. Hoe, and A. G. Nowatzyk. SimFlex: a fast, accurate, flexible full-system simulation framework for performance evaluation of server architecture. *SIGMETRICS Performance Evaluation Review*, 31(4):31- 35, April 2004.

[11] M. Horowitz. Scaling, power and the future of CMOS. In *20th International Conference on VLSI Design*, page 23, January 2007.

[12] 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 *IEEE International Symposium on Networks-on-Chip*, pages 124- 133, 2009.

[13] P. Koka, M. McCracken, H. Schwetman, X. Zheng, R. Ho, and A. Krishnamoorthy. Silicon-photonic network architectures for scalable, power-efficient multi-chip systems. In *37th Annual International Symposium on Computer Architecture*, pages 117- 128, 2010.

[14] A. Krishnamoorthy, R. Ho, X. Zheng, H. Schwetman, J. Lexau, P. Koka, G. Li, I. Shubin, and J. Cunningham. Computer systems based on silicon photonic interconnects. *Proceedings of the IEEE*, 97(7):1337 - 1361, July 2009.

[15] B. Lee, F. Doany, S. Assefa, W. Green, M. Yang, C. Schow, C. Jahnes, S. Zhang, J. Singer, V. Kopp, J. Kash, and Y. Vlasov. 20μm-pitch eight-channel monolithic fiber array coupling 160 Gb/s/channel to silicon nanophotonic chip. In *Conference on Optical Fiber Communications and National Fiber Optic Engineers Conference (OFC/NFOEC)*, pages 1 - 3, March 2010.

[16] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi. McPAT: an integrated power, area, and timing modeling framework for multicore and manycore architectures. In *42nd Annual International Symposium on Microarchitecture*, 2009.

[17] R. Merritt. ARM CTO: Power surge could create dark silicon. http://www.eetimes.com/electronics-news/4085396/ARM-CTO-power-surge-could-create-dark-silicon-, October 2009.

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

[19] Y. Pan, Y. Demir, N. Hardavellas, J. Kim, and G. Memik. Exploring benefits and designs of optically connected disintegrated processor architecture. *Workshop on the Interaction between Nanophotonic Devices and Systems (in conjunction. with MICRO-43)*, 2010.

[20] Y. Pan, J. Kim, and G. Memik. FeatherWeight: low-cost optical arbitration with QoS support. In *44th IEEE/ACM Annual International Symposium on Microarchitecture*, pages 105- 116, 2011.

[21] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary. Firefly: Illuminating future network-on-chip with nanophotonics. In *36th International Symposium on Computer Architecture*, 2009.

[22] J. Poulton, R. Palmer, A. Fuller, T. Greer, J. Eyles, W. Dally, and M. Horowitz. A 14-mW 6.25-Gb/s transceiver in 90-nm CMOS. *IEEE Journal of Solid-State Circuits*, 42(12):2745- 2757, 2007.

[23] B. M. Rogers, A. Krishna, G. B. Bell, K. Vu, X. Jiang, and Y. Solihin. Scaling the bandwidth wall: challenges in and avenues for CMP scaling. In *36th Annual International Symposium on Computer Architecture*, pages 371- 382, 2009.

[24] P. Rosenfeld, E. Cooper-Balis, and B. Jacob. DRAMSIM 2: A cycle accurate memory system simulator. *Computer Architecture Letters*, 10(1):16- 19, 2011.

[25] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan. Temperature-aware microarchitecture. In *30th Annual International Symposium on Computer Architecture*, pages 2- 13, 2003.

[26] C. Sun, C.-H. O. Chen, G. Kurian, L. Wei, J. Miller, A. Agarwal, L.-S. Peh, and V. Stojanovic. DSENT– a tool connecting emerging photonics with electronics for opto-electronic networks-on-chip modeling. In *6th International Symposium on Networks-on-Chip*, 2012.

[27] Y. Tamir and G. Frazier. Dynamically-allocated multi-queue buffers for VLSI communication switches. *IEEE Transactions on Computers*, pages 725- 737, 1992.

[28] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. G. Beausoleil, and J. H. Ahn. Corona: system implications of emerging nanophotonic technology. In *35th Annual International Symposium on Computer Architecture*, pages 153- 164, 2008.

[29] T. F. Wenisch, R. E. Wunderlich, M. Ferdman, A. Ailamaki, B. Falsafi, and J. C. Hoe. SimFlex: statistical sampling of computer system simulation. *IEEE Micro*, 26(4):18- 31, July-August 2006.

[30] M. Yang. A comparison of using icepak and flotherm in electronic cooling. In *7th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM)*, Vol 1, 2000.

[31] X. Zheng, J. E. Cunningham, I. Shubin, J. Simons, M. Asghari, D. Feng, H. Lei, D. Zheng, H. Liang, C. chih Kung, J. Luff, T. Sze, D. Cohen, and A. V. Krishnamoorthy. Optical proximity communication using reflective mirrors. *Optics Express*, 16(19), Sept. 2008.

[32] A. Zilkie, B. Bijlani, P. Seddighian, D. C. Lee, W. Qian, J. Fong, R. Shafiiha, D. Feng, B. Luff, X. Zheng, J. Cunningham, A. V. Krishnamoorthy, and M. Asghari. High-efficiency hybrid III-V/Si external cavity DBR laser for 3μm SOI waveguides. In *9th IEEE International Conference on Group IV Photonics (GFP)*, 2012.