Appl Phys A (2009) 95: 989–997
DOI 10.1007/s00339-009-5109-2

Applied Physics A
Materials Science & Processing

Devices and architectures for photonic chip-scale integration

J. Ahn · M. Fiorentino · R.G. Beausoleil · N. Binkert · A. Davis · D. Fattal · N.P. Jouppi · M. McLaren · C.M. Santori · R.S. Schreiber · S.M. Spillane · D. Vantrease · Q. Xu

Received: 29 August 2008 / Accepted: 16 December 2008 / Published online: 20 February 2009
© Springer-Verlag 2009

Abstract Silicon nanophotonics holds the promise of dramatically advancing the state of the art in computing by enabling parallel architectures that combine unprecedented performance and ease of use with affordable power consumption. This paper presents a design study for a many-core architecture called Corona which utilizes dense wavelength division multiplexing (DWDM) for on- and off-chip communication together with the devices which will be needed to implement such a communication infrastructure.

PACS 42.82.-m · 42.82.Ds · 42.82.Bq

1 Introduction

The sustained performance improvement of computational devices over the last 30 years has been supported by improved integrated circuit technology and the attendant decrease in transistor size and increase in transistor count per die. The industry focus on utilizing these benefits to improve single-thread performance led to increasingly complex computational cores and increased clock frequencies. This single-thread performance focus stumbled in the transition to a 90-nm process due to power, thermal, and associated cost problems. The merchant semiconductor industry then went through an architectural reset to redefine performance based on parallelism as it switched gears to multi-threaded and multi-core processor devices. The ITRS Semiconductor roadmap [1] predicts that, in the next 15 years, feature sizes will shrink from 40-nm to the sub-10-nm regime. This process technology improvement is expected to result in scaling the number of cores and threads per device rather than increased speed or complexity of an individual core. The validity of this prediction is in evidence as vendors are currently building chips with multiple relatively simpler cores with little or no increase in clock speed. It has been predicted [2] that this trend will continue and lead to devices with tens or hundreds of cores and up to thousands of threads running highly parallel codes. In order to achieve their performance potential, cores will need to access data stored in local and distant caches as well as in off-chip main memory. Programming these new devices efficiently is a significant problem. The programming problem will be simplified if the differences in memory bandwidth and latency between local and remote locations can be minimized: in this situation the programmer will not need to be concerned with data and program locality. The best network topology to achieve this architecture is a crossbar in which each core can be directly connected to any other core at any given time. In a purely electronic system achieving this high level of interconnection is unfeasible because of power considerations. The energy to transport a bit in an electrical network depends on distance and bandwidth is dependent on clock frequency, the number of wires in the communication channel, and the number of bits that can be transported simultaneously on each wire. All these factors result in power requirements that scale super-linearly with an increase in bandwidth.

Off-chip bandwidth is also limited by the number of package pins which is not predicted to increase significantly. The scaling problems associated with off-chip crossbars have led to the use of multi-hop network topologies, e.g., mesh, fat tree, folded Clos, etc., which mitigate the power

J. Ahn · M. Fiorentino (✉) · R.G. Beausoleil · N. Binkert · A. Davis · D. Fattal · N.P. Jouppi · M. McLaren · C.M. Santori · R.S. Schreiber · S.M. Spillane · D. Vantrease · Q. Xu
HP Laboratories, 1501 Page Mill Rd., MS 1123, Palo Alto, CA 94304, USA
e-mail: marco.fiorentino@hp.com

problem but do not adequately solve the bandwidth and latency issues. This motivates the pursuit of nonelectrical interconnect options.

This paper presents the Corona many-core architecture which utilizes a CMOS-compatible silicon photonic crossbar for intra- and inter-chip communication. The design assumptions for Corona are based on the ITRS [1] 17-nm technology node that is planned to be in production by 2017. Corona provides significant power savings, performance improvements, and improved programmability when compared with competing all-electrical solutions. Other proposals for on-chip nanophotonic interconnects have appeared recently, but most simply replace the long "global" wires with a bus [3] or a circuit-switched network [4]. Corona improves on the latency and bandwidth characteristics of these alternatives by more efficiently using the inherent advantages of silicon nanophotonics.

The remainder of the paper is organized as follows: Sect. 2 is dedicated to a description of the overall architecture and the network requirements; Sect. 3 describes the nanophotonic crossbar; Sect. 4 describes the performance of the Corona chip in terms of power consumption and simulated compute performance on a set of multi-threaded benchmarks.

2 The Corona architecture

Figure 1 shows a block diagram of Corona. Each 4-core cluster shares a memory controller for access to off-chip main memory. Multiple clusters communicate with each other through an on-chip optical interconnect network. Similar to the Intel Larrabee [5], the Corona cores are simple in-order, wide SIMD (Single Instruction Multiple Data) processing units. 3D die stacking technology allows the cluster tiles to be stacked on top of their associated L2 caches and memory controller tiles, an approach that is also used in the Intel Polaris [6] prototype. In a 17-nm technology, Corona will consist of 64 clusters, 256 cores, and provide a 10 TFLOP peak performance with a clock frequency of 5 GHz. The remaining problem to be solved is how to provide enough bandwidth between the 64 cores and 64 off-stack memory blocks to sustain this peak performance capability at a reasonable power consumption level.

One byte per FLOP for both on- and off-chip communication has been highly desirable in the high performance computing world [7]. However, providing this bandwidth (10 TB/s) is problematic in the all-electrical domain. For on-chip communication, full-swing long wires consume significant power, and low-swing wires are slow [8]. As a result, ring- or mesh-based topologies have been suggested for on-chip interconnection networks due to their short hop distances [5, 6]. In addition to the power issue, off-chip communication bandwidth is also limited by pin count and per-pin bandwidth limitation. Currently the best reported chip-to-chip transceiver technology uses 2.2 mW/Gb/s [9]. This implies that more than 170 W will be consumed to just provide off-chip connectivity for Corona's needs. Numerous software tactics may mitigate the bandwidth demand, but this adds both programmer or compiler complexity, and the techniques are often application specific. In order to solve this performance limiting bandwidth problem, a new technology approach appears necessary.

3 The nanophotonic crossbar

Optical interconnects have been successful for long-distance data transmission and are becoming increasingly cost-effective for shorter distances. They are therefore an obvious candidate to investigate as a solution for Corona's communication needs. A compelling advantage of optical interconnects is the ability to increase the bandwidth of a single physical channel by using wavelength division multiplexing

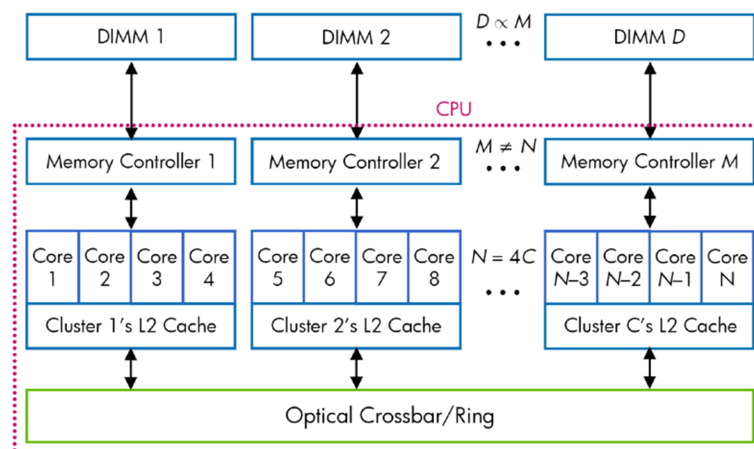


Fig. 1 Schematic of the basic architecture for a 256-core processor. The cores are divided into clusters that share an L2 cache and control access to a particular unit of memory

(DWDM). Given the power overheads incurred by serialization and deserialization (SERDES), it is unlikely that on-chip interconnects will be driven at rates higher than twice the clock speed. Since active power is linear with clock frequency, clock speeds are not likely to exceed 5 GHz in the next decade [1]. Therefore to achieve a 20 TB/s bandwidth, one would need 16,000 physical channels running at 10 Gb/s. A 16,000 channel electrical crossbar would be infeasible from both a power and cost perspective. Similar limitations would apply for a coarse wavelength division multiplexing (CWDM) optical interconnect. A DWDM optical approach remains, but it presents significant implementation challenges. If 64 wavelengths can be transmitted over any waveguide, the 10 TB/s photonic bandwidth target can be achieved with 250 waveguides that can be easily accommodated in one layer.

The feasibility of such on- and off-chip photonics will require that these integrated photonic circuits be compatible with conventional CMOS fabrication processes. Recent progress in silicon photonics has shown that most of the devices needed for such networks are indeed possible. In addition, when economies of scale are taken into account, it is reasonable to believe that Si-based photonic integrated circuits can be produced at costs that are competitive with the alternative solutions. Furthermore the necessary 3D chip stacking technologies are likely to mature by 2017 [10].

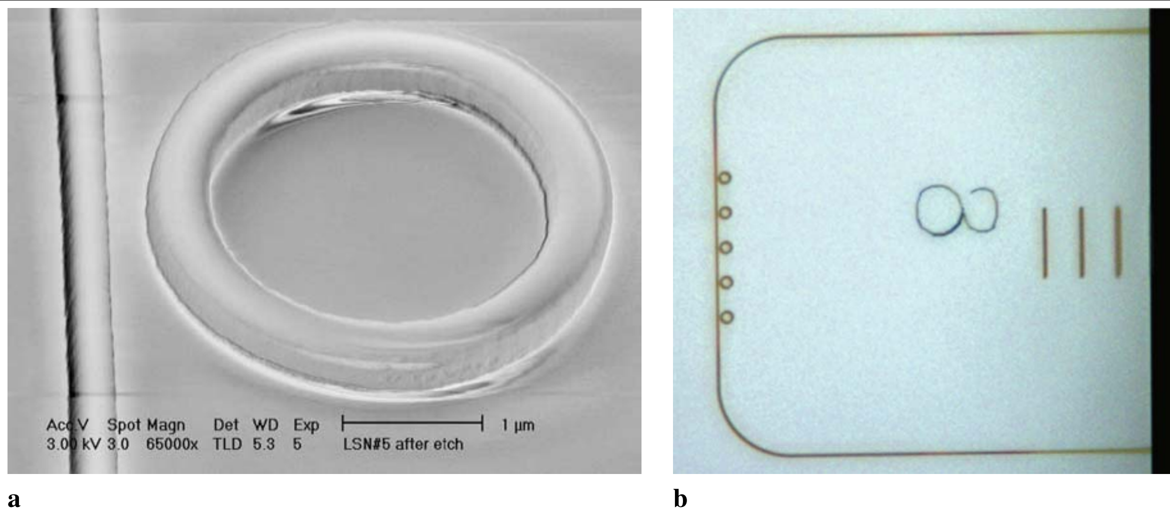
A silicon photonic integrated circuit (PIC) requires a low-loss waveguide. Silicon-on-insulator (SOI) waveguides losses as low as 0.2 dB/cm [11] have been measured, and will not need much improvement. However, the commercially available SOI wafers used for this purpose are custom-made and expensive. To satisfy light confinement requirements, a thick buried oxide layer is necessary, and this layer thermally isolates the devices which exacerbates thermal management problems in a PIC. In the long term, it will be important to develop a means for creating nanophotonic components using pure silicon wafers, which provide high thermal conductivity at low cost.

Another key component for a Si-PIC is a suitable light source. Silicon light sources are very hard to build because of the material's indirect bandgap. Therefore a III–V laser will need to be used to generate light either off-chip or as an on-chip hybrid laser. Recently multi-wavelength lasers with precisely controlled frequency spacings have been built. One advantage of these lasers is that if only one of the frequency channels is servo-locked to an on-chip standard cavity, then all of the other frequency modes will track the controlled mode. One possible approach is the Fabry–Perot comb laser based on quantum dots [12], which has already been used to demonstrate a bit-error-rate of 10^{-13} at 10 Gb/s over ten longitudinal modes [13, 14]. Another possible approach is the mode-locked hybrid Si/III–V evanescent laser [15], which uses a silicon-waveguide laser cavity wafer-bonded to a III–V gain medium. Any temperature change in the environment will cause approximately the same refractive index shift in the laser cavity and the silicon waveguides and resonators that form the DWDM network which simplifies locking. Overall a wavelength locking scheme that is robust against temperature changes is one of the main implementation challenges on which we are currently working.

To implement a DWDM network it is also necessary to build frequency selective modulators and detectors. A DWDM modulator can, in-principle, be built using wavelength-independent modulators (such as Mach–Zehnder electrooptic modulators [16] and quantum-well materials for electro-absorption modulators [17, 18]) and add-drop filters. This solution is not easily implementable given the large area of nonresonant add-drop filters such as array waveguide gratings. A combined solution is possible for detectors, but it suffers from similar problems. An alternative solution is to use resonant elements to multiplex and demultiplex the signals. This solution is more complex to implement because resonant elements are more sensitive to environmental changes such as temperature and require very strict fabrication tolerances.

The Corona effort has therefore been based on the silicon microring resonator because of its small size, high quality factor Q, transparency to off-resonance light, and small intrinsic reflections. Using injected charge, the refractive index of the microring can be changed to blue-shift the fundamental frequency of the cavity. This mechanism can be used to move the ring into or out of resonance with an incident light field thus providing a mechanism for electro-optic modulation [19–21] in an on-off keying scheme. When not modulating, the rings are in the "OFF" position and do not interfere with light transmission. A slower tuning mechanism is provided by temperature: increasing the ring temperature red-shifts the frequency, and this mechanism can be used to track slow changes in the laser frequency. When a small amount of absorbing material such as Ge is incorporated into the ring, the resonator can be used as a frequency-selective detector. Additional functionality can be obtained by using a variable-wavelength add-drop filter [22, 23] as a switch. A ring with a radius of 1.5 μm and an intrinsic quality factor Q of 18,000 has been demonstrated recently by our group. This quality factor only takes into account intrinsic losses in the ring; when losses caused by the ring coupler are added, the result is a "loaded" Q of 9,000. Effective mode volumes for these rings are around 1.0 μm^3 [24]. The measured quality factor is close to the maximum achievable Q given Si waveguide bending losses. In Fig. 2(b), cascaded silicon microring resonators are shown. These can be used as a modulator or filter bank in a nanophotonic network.

There are many obvious advantages in building small rings. First, smaller components make possible a larger amount of integration and greater on chip-functionality.



a **b**

Fig. 2 (**a**) An SEM picture with 40°-titled view of a microring resonator with a 1.5-μm radius coupled to a waveguide with an optimized (reduced) width. (**b**) A microscope picture of cascaded microring resonators coupled to a U-shaped waveguide at the edge of the chip

In addition, smaller components have smaller capacitance thus requiring less power to switch. This is also true for the temperature-tuning mechanism where less thermal mass corresponds to less power required to stabilize temperature. The total bandwidth of a microring-based DWDM modulation system [20] is limited by the free spectral range (FSR) of the microring resonator, which is inversely proportional to the circumference of the ring. A smaller microring modulator has a larger FSR, which can therefore accommodate more wavelength channels and have higher aggregate data bandwidth. A 1.5-μm-radius ring with a coupled Q of 9,000 has been demonstrated in our lab [24]. This provides an FSR of about 10 THz, and a filter bandwidth of about 10 GHz, which is nearly ideal for the Corona interconnect architecture.

A schematic of the photonic crossbar and electronic circuit is shown in Fig. 3. The Corona design is based on several stacked layers dedicated to specific functions. Each layer comprises 64 tiles that are vertically associated in the die stack. The top layer consists of 64 cluster tiles, each of which is comprised of four processor cores and their associated private L1 caches. Each second layer tile consists of an L2 cache, main memory controller, a directory used to support L2 cache coherence, and a communication hub used to direct communications to and from the appropriate subsystem in the tile. The 64 processor/memory tiles in the top two layers constitute the nodes of the optical network which is implemented in the next two lower layers. The third layer consists of analog electronics which interface the top two digital layers to the optical waveguides and resonance rings on the bottom layer. The main portion of the network comprises 256 data/control waveguides each carrying 64 wavelengths. The DWDM channels are separated by 80 GHz in order to allow transmission at 10 Gb/s on each channel and reduce losses for off-line modulators. This choice does not conform to the ITU channel spacing, but it is chosen to optimize the bandwidth density. We feel that because of the wavelength that we are using (around 1310 nm) and the type of application (short-haul datacom as opposed to long-haul telecom) conforming to the ITU grid would have negligible effect on the availability and cost of the optoelectronic components while putting severe constraints on the available bandwidth. The waveguides are grouped in 64 bundles of four waveguides, and each bundle starts at one node (with splitters that inject optical power from a power waveguide), snakes its way past all the tiles, and returns to the original node where it is terminated by 64 resonant detectors. The bundle can be used to communicate with the node where the bundle terminates by any of the other nodes; this is accomplished by using the appropriate set of modulators to write data onto the bundle. At any time 64 nodes can be transmitting each using 256 channels (four waveguides with 64 wavelengths each) that can be modulated at 10 Gb/s for an aggregated bandwidth of 20 TB/s. When error correction and other overheads are taken into account, this will provide a usable data bandwidth in excess of the target of 10 TB/s. Note that light and data always flow in the same direction (clockwise in the schematic of Fig. 3). Because multiple nodes may want to communicate on the same bundle, arbitration is required. To achieve this we devised a scheme for all-optical arbitration that is based on an optical token that is passed around a dedicated arbitration ring (shown in Fig. 3). The arbitration scheme is discussed in [26]. In addition to arbitration, there are a number of waveguides dedicated to a broadcast channel which is used to coordinate L2 cache coherence. Finally each node has four dedicated waveguides (two for input and two for output) dedicated to communication with external optically connected memory

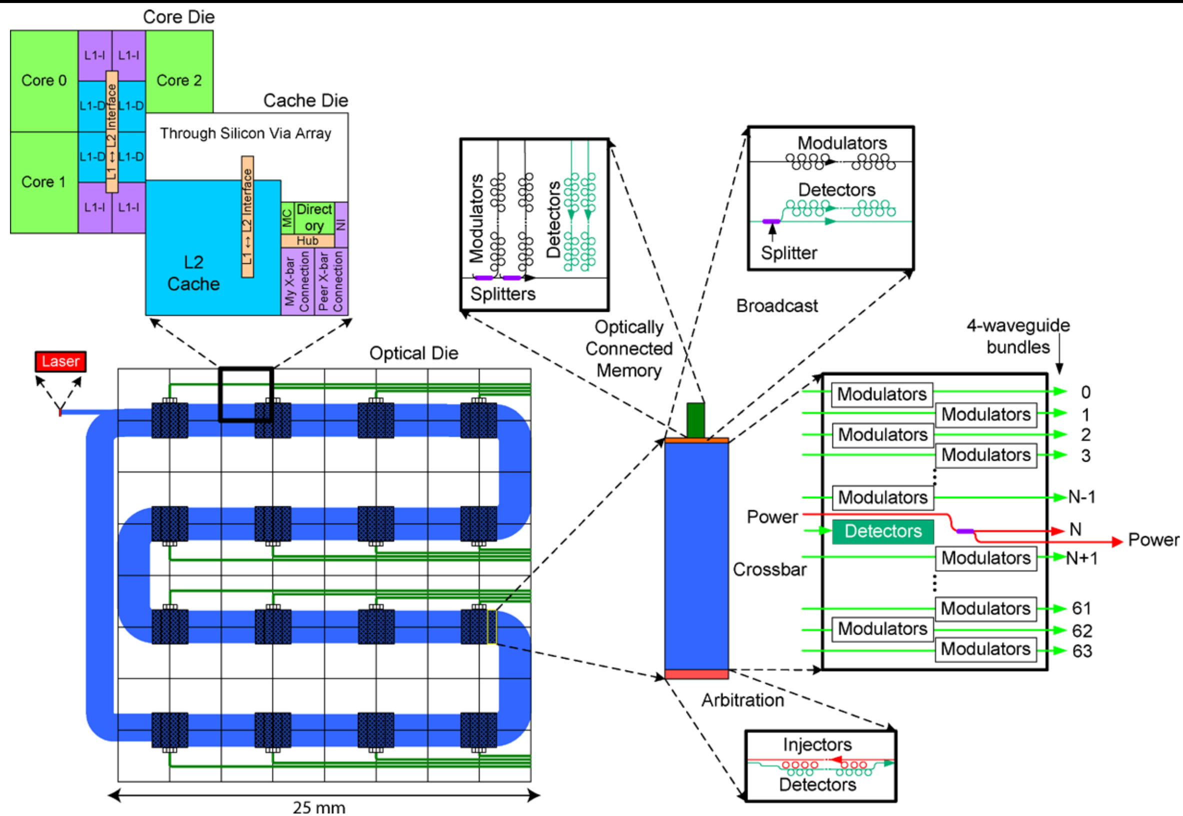


Fig. 3 The nanophotonic interconnect die for a 256-core chip operating at 5 GHz. Subsystems illustrated here include: the 64 cluster tiles on the processor and memory dies; 270 identical parallel ridge waveguides allocated to cache line transfers and control messages (*blue*), and additional waveguides providing off-chip I/O (*green*); an expanded view of a photonic data/control block, an arbitration block, a broadcast block, and an off-chip communication hub; and laser power distribution

(OCM). These waveguides carry 64 wavelengths for a total bandwidth of 10 TB/s in each direction.

4 Corona performance

4.1 Power consumption

Power consumption is of paramount importance for Corona. There are several key aspects that contribute to the overall power consumption of the optical interconnect: power used to generate the light, electrical power used to modulate the rings, and the power needed to keep the rings on resonance.

To determine how much laser power is needed, a link loss approach was used. The link loss is defined as the ratio between the number of electrons that are injected in the laser and the number of electrons that come out of the detector at the end of the link. In the case of the on-chip network, this budget includes: the efficiency in the conversion of electrons to photons in the laser (5 dB in 2017, corresponding to a 30% wallplug efficiency), transmission losses (a total of 11 dB including: 1 dB coupling losses onto the chip, 3.5 dB waveguide propagation losses, 6.4 scattering losses from off-resonance rings, and 0.1 dB losses from splitters), and the conversion efficiency of photons into electrons (3 dB). The loss numbers were calculated using literature results (e.g., 1 dB/cm loss for single-mode waveguides and 0.3 dB/cm for multi-mode waveguides [25]) and numerical simulations (e.g., for the ring scattering losses). The total link loss budget is 19 dB. To estimate the number of electrons needed at the output of the photodiode to flip a transistor, we assume that photodiodes with a capacitance of $\simeq$10 fF will be available. This capacitance target can be reasonably achieved if proper scaling is applied to existing results [27]. Low-capacitance detectors will allow one to detect signals with as little as 1000 electrons/bit without significant power overhead. By backing up the losses, one can calculate the power necessary to generate laser light for the Corona on-chip network to be 2.6 W (this corresponds to 0.8 W of laser power). Similar estimates for the off-chip network give an estimated power consumption of 0.4 W for the off-chip network.

The other main contribution to power dissipation comes from the rings. The rings dissipate power in three different ways: fabrication error trimming, resonance frequency biasing, and direct data modulation. Because of fabrication

imperfections, each ring will have a resonance frequency that is slightly different from the design goal, and must be "trimmed" into the correct spectral location. The ring will also need to be "biased" into position when ready to modulate. There are two schemes to fine-tune the rings: carrier injection to blue-shift the resonance and thermal heating for red-shifting. In the worst-case scenario, 190 μW/nm are needed to red-shift a 3-μm ring through heating, and 130 μW/nm are needed to blue-shift via current injection. On average, a wavelength offset of 0.14 nm will need to be corrected for the expected fabrication tolerances at the 17-nm technology node. This corresponds to an average power dissipation of 17 and 26 μW for blue- and red-shifting, respectively. When multiplied by the total number of rings on the circuit (approximately 1.1×10^6), this gives us a total power consumption of $\simeq$26 W.

To modulate a 3-μm ring one will need a charge of 3×10^{-14} C/pulse at 5 GHz, or 30 fJ at 1 V. This number is calculated assuming that the ring is detuned 40 GHz from resonance to obtain an extinction ratio of 10 dB or greater. This corresponds to a raw dissipated power of 150 μW per online ring. The modulator driver will necessarily dissipate electrical power, since the modulator acts as a capacitive load with a 10 μA leakage current in the "on" state and has a peak current of 1 mA during the transition. A careful co-design of the voltage drivers is of paramount importance to achieve the power consumption targets stated here. We believe that it is important to develop *analog* CMOS drivers to reduce the electronic overhead by a factor of 3. With that assumption, the power dissipated to modulate each ring is $\simeq$0.5 mW at the 17-nm technology node. Considering that there are $\simeq$16000 active rings in the on-chip network and $\simeq$16000 in the off-chip network (including the ones on the optically connected memories), the total power needed for modulation is estimated to be $\simeq$16 W. The total power needed to bias the actively modulated rings into position can be calculated to be $\simeq$3 W.

Aggregating the various power components results in a total power of $\simeq$48 W. This should be compared with the estimated power consumption of the processor and memory nodes that approaches 200 W. The data in Table 1 show the expected evolution of power consumption at various technology nodes using network solutions similar to Corona.

Table 1 Projected chip interconnect performance at various technology nodes. The power consumption of the interconnect includes lasers and modulators, as well as the power needed to keep resonant detectors and modulators locked

	Technology node (nm)			
	40	28	17	14
Cluster/chip	4	16	64	64
Cores/cluster	4	4	4	16
TFLOP/s	0.64	2.56	10.24	40.96
On-chip interconnect				
Bandwidth (TB/s)	1.3	5.1	20.5	81.9
Power (W)	3.8	18.2	38.4	118.4
Energy/bit (fJ)	367	445	235	181
Off-chip interconnect				
Bandwidth (TB/s)	1.3	5.1	20.5	81.9
Power (W)	1.8	4.3	8.9	27.5
Energy/bit (fJ)	177	105	54	42

Table 2 Description of synthetic workloads

Workload	Description
Uniform	From cluster (i, j) to a random cluster
Hot Spot	From cluster (i, j) to a fixed cluster
Tornado	From cluster (i, j) to cluster $((i + \lfloor k/2 \rfloor - 1)\%k,\ (j + \lfloor k/2 \rfloor - 1)\%k)$, where k the radix of the network
Transpose	From cluster (i, j) to cluster (j, i)

4.2 System performance

A simulation of the Corona system is conducted to show how the high bandwidth and low communication energy enabled by the optical interconnect are translated into system performance and efficiency gains in many-core applications. A combination of synthetic (described in Table 2) and realistic (the SPLASH-2 benchmark suite [28]) workloads is used for the evaluation. The COTSon [29] and M5 [30] framework were used for the simulations.

The main aim is to understand the performance implications of the optical interconnects between clusters and to the off-chip memory. The simulator was configured for three network options:

- XBar—An optical crossbar [26] with bisection bandwidth of 20.48 TB/s and maximum signal propagation time of 8 clocks.
- HMesh—An electrical 2D mesh [26] with bisection bandwidth 1.28 TB/s and per hop signal latency (including forwarding and signal propagation time) of 5 clocks.
- LMesh—An electrical 2D mesh [26] with bisection bandwidth 0.64 TB/s and per hop signal latency (including forwarding and signal propagation time) of 5 clocks.

The results of the previous section are used to estimate that the optical crossbar consumes a continuous power of 29 W for the Xbar since many components of the optical system power are fixed (e.g., laser, ring trimming, etc.). For the electrical mesh, an energy of 196 pJ per transaction per hop, including router overhead, is used based on an aggressive use of low swing busses and ignoring any mesh leakage power.

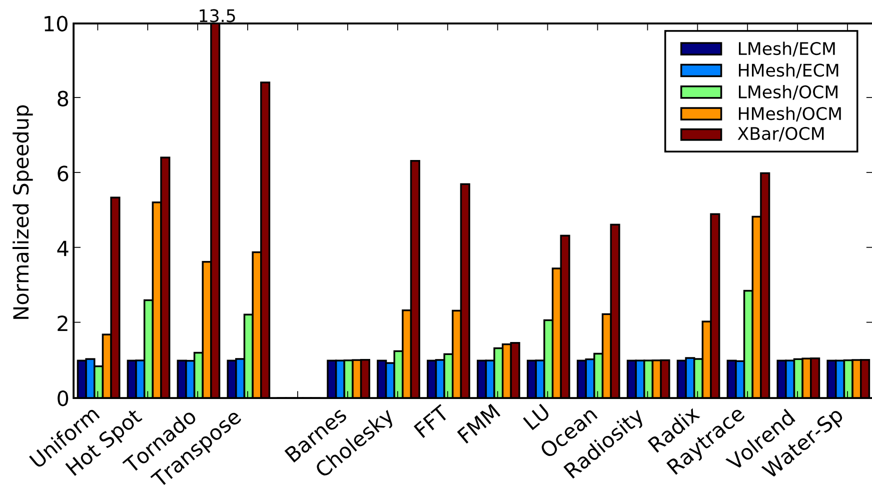


Fig. 4 The performance of synthetic and realistic workloads. The synthetic workloads are placed on the *left side* and explained in Table 2. The SPLASH-2 benchmark suite [28] is used for the realistic workloads, which are shown on the *right side*. OCM provides substantial performance gain when the memory bandwidth demands of workloads are not low, and the optical crossbar provides additional benefits on most of these workloads

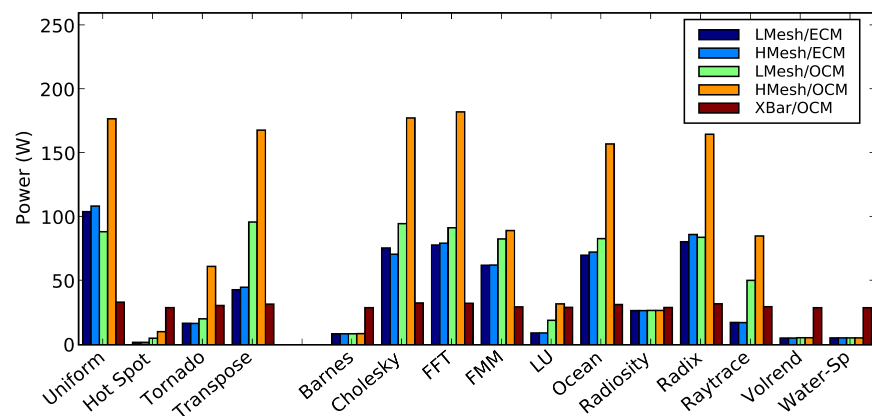


Fig. 5 The on-chip network power on synthetic and realistic workloads. When the memory bandwidth demands of workloads are not low, the electrically connected systems (ECM, LMesh, and HMesh) consume much higher power than the Xbar/OCM system

Two memory interconnects were also simulated, the OCM interconnect plus an electrical interconnect:

– OCM—Optically connected memory; main memory bandwidth is 10.24 TB/s, memory latency is 20 ns.
– ECM—Electrically connected memory; main memory bandwidth is 0.96 TB/s, memory latency is 20 ns.

The electrical memory interconnect is based on the ITRS [1] roadmap, which indicates that it will be impossible to implement an ECM with performance equivalent to the proposed OCM.

Five combinations in total are simulated: XBar/OCM (Corona), HMesh/OCM, LMesh/OCM, HMesh/ECM, and LMesh/ ECM in order to highlight the performance gain due to faster memory and interconnect per benchmark. Further details of the experimental setups are described in [26].

Figure 4 shows the performance of synthetic and realistic workloads for the five system configurations relative to the electrically connected LMesh/ECM system. When the memory bandwidth is low, the faster mesh has little impact. With a fast OCM, substantial performance gain is observed over ECM systems when the fast mesh or the crossbar interconnect is used, while much less gain is observed with the low performance mesh. Most of the performance gain made possible by OCM is realized only if the crossbar interconnect is used. Hot Spot is the exception: its performance is memory bandwidth limited since all memory traffic is sent through a single cluster, and the interconnect is not the bottleneck. Overall, by moving to an OCM from an ECM in systems with an HMesh, a geometric mean speedup of 3.28 is achieved. Adding the photonic crossbar can provide a further speedup of 2.36 on the synthetic benchmarks.

For the SPLASH-2 applications, in five cases (Barnes, Radiosity, Volrend, FMM, and Water-Sp) the LMesh/ECM system is adequate. These applications perform well due to their low cache-miss rates and consequently low main memory bandwidth demands. The remaining applications are memory-bandwidth limited on ECM-based systems. For Cholesky, FFT, Ocean, and Radix, fast memory provides considerable benefits, which are realized only with the fast crossbar. LU and Raytrace are similar to Hot Spot: while OCM provides most of the speedup, some additional benefit is derived from the use of the fast crossbar. Overall, replacing an ECM with an OCM in a system using an HMesh can provide a geometric mean speedup of 1.80. Adding the pho-

tonic crossbar can provide a further speedup of 1.44 on the SPLASH-2 applications.

Figure 5 shows the on-chip network power. Note that the Xbar continually uses power since light is being transmitted even though it is not being used. However, the electronic meshes also have leakage power, which is not included in Fig. 5. For main-memory intensive workloads (Uniform, Transpose, Cholesky, FFT, FMM, Ocean, and Radix), electronic meshes provide reduced performance (Fig. 4) at prohibitive power levels of 100 W or more (Fig. 5). The on-chip network power is particularly higher on the HMesh/OCM configuration since the high memory bandwidth made available by the optically connected memory stresses the on-chip network bandwidth of the HMesh, leading to much higher power consumption.

5 Conclusions

The implementation of a chip like Corona would require a shift in the silicon industry and will need to overcome several important technological hurdles. While individual components for Corona have been demonstrated, the massive integration of photonic circuits advocated here still remains to be demonstrated. To achieve this level of integration, a road-map for a Moore's law for photonics will be needed. In addition to refining the design of photonic devices and defining a set of photonic design rules, a directed effort to develop effective ways to co-design photonic and electronic devices with the target of reducing overall power consumption is required. A significant effort will also be needed to study the network implications of crosstalk (from adjacent DWDM channels and polarization conversion) as well as coding and modulation techniques that can be used to limit its detrimental effects.

Corona represents a feasibility blueprint for the implementation of an on-chip photonic interconnect and the architectural implications of such interconnect. The main advantage of a photonic interconnect is improved on- and off-chip bandwidth at affordable power levels, which cannot be achieved by an all-electrical interconnect. This enables the implementation of a many-core processor for which the byte-per-FLOP ratio (i.e., the ratio of the communication bandwidth to the compute bandwidth) is 1 both for the on-chip global interconnect and the off-chip memory links. This in turn enables shared-memory parallel architectures that are easier to program. In addition, having photonics on the chip enables new implementations of broadcast busses and arbitration protocols that further improve performance by reducing latency.

References

1. http://www.itrs.net/
2. K. Asanovic et al., The landscape of parallel computing research: a view from Berkeley. Technical Report UCB/EECS-2006-183, EECS Department, University of California, Berkeley, December 2006
3. N. Kirman, M. Kirman, R.K. Dokania, J. Martinez, A.B. Apsel, M.A. Watkins, D.H. Albonesi, Optical technology in future bus-based multicore designs: opportunities and challenges. IEEE Micro **27**, 56–66 (2007)
4. K. Bergman, L. Carloni, Power efficient photonic networks on-chip, in *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 6898 (2008), p. 689813
5. L. Seiler et al., Larrabee: a many-core x86 architecture for visual computing, in *SIGGRAPH*, August 2008
6. S. Vangal et al., An 80 Tile 1.28 TFLOPs Network-on-Chip in 65 nm CMOS, in *ISSCC*, February 2006
7. D.E. Atkins, K.K. Droegemeier, S.I. Feldman, H. Garcia-Molina, M.L. Klein, D.G. Messerschmitt, P. Messina, J.P. Ostriker, M.H. Wright, Revolutionizing science and engineering through cyber-infrastructure. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, January 2003
8. R. Ho, On-chip wires: scaling and efficiency. PhD thesis, Stanford University, 2003
9. R. Palmer et al., A 14 mW 6.25 Gb/s transceiver in 90 nm CMOS for serial chip-to-chip communications, in *ISSCC*, February 2007
10. B. Black et al., Die stacking (3D) Microarchitecture, in *Proceedings of the 39th International Symposium on Microarchitecture*, December 2006
11. A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, M.J. Paniccia, A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor. Nature **427**, 615–618 (2004)
12. A. Kovsh, I. Krestnikov, D. Livshits, S. Mikhrin, J. Weimert, A. Zhukov, Quantum dot laser with 75 nm broad spectrum of emission. Opt. Lett. **32**, 793–795 (2007)
13. A. Gubenko, I. Krestnikov, D. Livshtis, S. Mikhrin, A. Kovsh, L. West, C. Bornholdt, N. Grote, A. Zhukov, Error-free 10 Gbit/s transmission using individual Fabry–Perot modes of low-noise quantum-dot laser. Electron. Lett. **43**, 1430–1431 (2007)
14. http://www.innolume.com/
15. B.R. Koch, A.W. Fang, O. Cohen, J.E. Bowers, Mode-locked silicon evanescent lasers. Opt. Express **15**, 11225–11233 (2007)
16. W.M. Green, M.J. Rooks, L. Sekaric, Y.A. Vlasov, Ultra-compact, low RF power, 10 Gb/s silicon Mach–Zehnder modulator. Opt. Express **15**, 17106–17113 (2007)
17. Y.-H. Kuo, Y.-K. Lee, Y. Ge, S. Ren, J.E. Roth, T.I. Kamins, D.A.B. Miller, J.S. Harris, Strong quantum-confined stark effect in germanium quantum-well structures on silicon. Nature **437**, 1334–1336 (2005)
18. J.E. Roth, O. Fidaner, R.K. Schaevitz, Y.-H. Kuo, T.I. Kamins, J.S. Harris, D.A.B. Miller, Optical modulator on silicon employing germanium quantum wells. Opt. Express **15**, 5851–5859 (2007)
19. Q. Xu, B. Schmidt, S. Pradhan, M. Lipson, Micrometre-scale silicon electro-optic modulator. Nature **435**, 325–327 (2005)
20. Q. Xu, B. Schmidt, J. Shakya, M. Lipson, Cascaded silicon micro-ring modulators for WDM optical interconnection. Opt. Express **14**, 9430–9435 (2006)
21. Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, M. Lipson, 12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators. Opt. Express **15**, 430–436 (2006)
22. S. Xiao, M.H. Khan, H. Shen, M. Qi, A highly compact third-order silicon microring add-drop filter with a very large free spectral range, a flat passband and a low delay dispersion. Opt. Express **15**, 14765–14771 (2007)
23. M.S. Nawrocka, T. Liu, X. Wang, R.R. Panepucci, Tunable silicon microring resonator with wide free spectral range. Appl. Phys. Lett. **89**, 071110 (2006)

24. Q. Xu, D. Fattal, R.G. Beausoleil, Silicon microring resonators with 1.5-μm radius. Opt. Express, **16**, 4309–4315 (2008)
25. M. Lipson, Guiding, modulating, and emitting light on silicon challenges and opportunities. J. Lightwave Technol. **23**, 4222 (2005)
26. D. Vantrease et al., Corona: system implications of emerging nanophotonic technology, in *International Symposium on Computer Architecture* (2008), pp. 153–164
27. T. Yin et al., 31 GHz Ge n-i-p waveguide photodetectors on silicon-on-insulator substrate. Opt. Express **15**, 13965 (2006)
28. S.C. Woo, M. Ohara, E. Torrie, J.P. Singh, A. Gupta, The SPLASH-2 programs: characterization and methodological considerations, in *International Symposium on Computer Architecture* (1995)
29. A. Falcon, P. Faraboschi, D. Ortega, Combining simulation and virtualization through dynamic sampling, in *ISPASS*, April 2007
30. N.L. Binkert, R.G. Dreslinski, L.R. Hsu, K.T. Lim, A.G. Saidi, S.K. Reinhardt, The M5 simulator: modeling networked systems. IEEE Micro **26**(4) (2006)