IEEE Asian Solid-State Circuits Conference
November 16-18, 2009 / Taipei, Taiwan

Circuits for silicon photonics on a "macrochip"

Ron Ho, Jon Lexau, Frankie Liu, Dinesh Patil, Robert Hopkins, Elad Alon[2], Nathaniel Pinckney, Philip Amberg, Xuezhe Zheng[1], John E. Cunningham[1], and Ashok V. Krishnamoorthy[1]
Sun Microsystems Laboratories, Menlo Park, CA, USA
[1]Sun Microsystems Chief Technology Office, San Diego, CA, USA
[2]Department of EECS, University of California, Berkeley, CA, USA
ron.ho@sun.com

***Abstract*—Recent advances in silicon photonics bring significant benefits to "macrochip" grids made of arrayed chips. Such configurations have global interconnects long enough to benefit from the high speed, low energy, and high bandwidth density of optics. In this paper we consider the constraints of large macrochip systems, and explore modulator drivers and photodetector receivers that match those constraints. We show measured results from a recent 90 nm testchip intended to mate with optical components.**

I. Introduction

Many researchers are developing silicon photonics technologies in order to bring optical communications onto a VLSI chip. Such an optical link uses a transmitter to turn an electrical bit stream into modulated optical energy, an optical waveguide or set of waveguides running from source to sink, and a receiver to convert optical energy back into electrical signals. In order to also share waveguides among different bitstreams using wavelength division multiplexing (WDM), the path would also need one or more wavelength "add" muxes as well as wavelength "drop" demuxes.

All of these components can be fabricated in a modern integrated circuit technology, although some optical devices would require process modifications from a base CMOS VLSI flow. These include adding Ge to create photodetectors, requiring a buried oxide layer similar to that in SOI wafers to confine light in silicon waveguides, and so on. Several such optical devices have been demonstrated in recent years [1-5].

The putative benefits of such an optical path over an electrical one include lower latency: the optical bits travel at the waveguide's speed of light and not at a speed set by *RC* time constants. Also, because waveguide optical losses are small, energy costs of optical paths are largely independent of distance, unlike electrical signaling energy costs that scale with wire length. Finally, optical paths with WDM in shared waveguides have a higher bandwidth density than electrical wires, improving interconnect routing and reducing hotspots.

However, a closer look at these benefits for a large-scale VLSI system shows that the use of optics provides truly compelling gains for commercial systems when the distance traveled by signals exceeds that of a single chip. Large "macrochips," (see Fig. 2) enabled by technologies that can stitch together smaller chips with extremely dense chip-to-chip input/output (I/O) paths, have global signals long enough to benefit from silicon photonics. In this paper we discuss these constraints and explore the types of circuit architectures motivated by such topologies. We also show some preliminary results from a research program aimed at building optical and electrical components for such macrochips.

This work is supported in part by the U.S. Defense Advanced Research Projects Agency under HR0011-08-09-0001.

II. Optical and Electrical Interconnect

In this discussion we assume a next-generation 32 nm high-performance CPU. That is, we envision a large chip (400-700 mm^2) running at a moderately high clock rate (2-5 GHz), for which interconnect energy, latency, and bandwidth density are all important.

Cross-chip interconnect on such a chip traditionally uses *RC* wires with periodic repeaters, at well-understood costs [6]. Energy per cycle using CMOS circuits is kCV^2, where C is total wire capacitance and k combines a 1.2x capacitance overhead for repeaters with a 0.25x term for switching activity factor. This energy cost is linear in wire length, and is about 75 fJ/mm/cycle for 1 V power supplies. The delay of repeated *RC* wires is a geometric mean of wire delay and gate delay, and is around 16x worse than the speed of light, or 100 ps/mm. Finally, the bandwidth density of repeated wires depends on pitch and the number of metal layers; using two wire layers per direction and wires 4x the minimum width, we can fit 1.5 global wires/μm of cross-sectional width.

Designers can certainly trade these characteristics against each other. Energy costs can be lowered by reducing voltage swing, to 30 fJ/mm/cycle at a 50 mV swing (scaling results from [7] to 1 V); or by using feed-forward equalization circuits to achieve 20 fJ/mm/cycle (scaling results from [8] to 1 V). Both methods also cut latency, but not as much as transmission lines do [9]. These techniques trade energy for bandwidth density: low-swing or equalized wires are differential to reduce noise effects, halving bandwidth density. Transmission lines often require more than 10 microns per microstrip line, and worse yet, require other nearby metal layers to be empty, dramatically lowering bandwidth density.

For designers to consider replacing electrical wires with optical waveguides, given integration and schedule risks, the improvement in performance metrics must be large. Over a 20 mm global route, a repeated wire will incur energy costs of 0.4-0.6 pJ/cycle and latency around 1.8 ns, or 3-9 clock cycles. However, even aggressive projections of silicon photonics anticipate only barely beating these energy targets [10].

978-1-4244-4434-2/09/$25.00 ©2009 IEEE

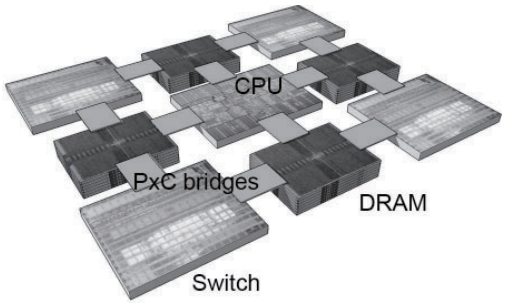

Figure 1: An example system using proximity communication, integrating into a single package a processor, DRAM stacks, and switch I/O chips (from [12]).

Also, the latency of a silicon photonics link would be at most 8 clock cycles faster, and less after accounting for electrical-optical and optical-electrical conversions. While this is not a trivial difference, architects employing multicore/multithread chip topologies can mitigate such global latencies through some added chip complexity. Finally, assuming a 5 μm waveguide pitch and the ability to multiplex 16 colors per waveguide gives optics a 4x bandwidth density advantage—a nice benefit, but not enough for a design win.

III. Large-scale "Macrochips"

A 20 mm global interconnect is not sufficiently long to show an obvious advantage for optics versus electronics. However, a 100 mm long on-chip wire would be: energy costs for this wire would nearly 10x higher, and latency costs more than 40 cycles longer, than those of an aggressive optical path. Of course, a 100 mm on-chip wire also does not exist on a VLSI chip.

A. Proximity communication

High-bandwidth chip-to-chip communication is traditionally overclocked and sent through solder balls and package and board traces. An alternative technology forms chip-to-chip connections by placing chips in close face-to-face proximity with each other. Metal pads on each chip pair up to form capacitors, which can be extremely dense (24 μm pitch) and low energy [11]. Chips need not be soldered together [13], so large arrays of reworkable chips can be created, and yield limits on multi-chip packages relaxed. Because these proximity connections are nearly as dense as on-chip wires, these arrays form virtual big chips that look like a monolithic silicon die. Fig. 1 shows a cartoon of an example 3x3 system, with a processor, DRAM stacks, and I/O switching chips. Spanning each pair of face-up functional "island" chips is a face-down "bridge" chip that uses proximity communication to send data into and out of islands. If each chip were 20-25 mm on a side, any corner-to-corner communication across the system would run along more than 100 mm of on-chip wires, along with many intermediate proximity communication hops.

B. Optically-enabled macrochips

Using optics improves the performance of long interconnects in large proximity communication arrays and enables further scaling up in system size. However, routing optical signals across the many face-to-face chip pairings in Fig. 1 causes optical loss at each interface and hence worse energy efficiency [14]. Instead, we envision an optically-enabled "macrochip" system, portrayed in Fig. 2, that uses a silicon wafer alignment lattice to register island chips (CPUs, DRAM stacks, *etc.*). The lattice also contains optical waveguides in both lateral directions. Bridges with silicon photonic devices overhang each island chip and the lattice, providing for electrical microsolder to the island and optical proximity coupling to the waveguides. Many more details, including overviews of optical devices, loss budgets, and a sample architecture, are shown in [10].

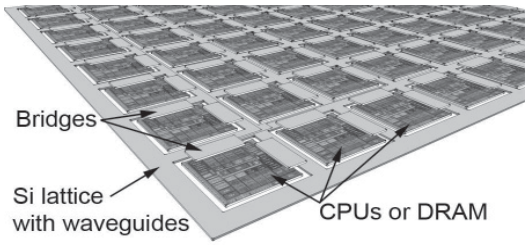

Figure 2: An 8x8 macrochip, with face-up island chips registered in a lattice containing waveguides. Face-down bridges connect the islands to waveguides in the lattice. Many more details are given in [10].

IV. Circuit constraints and topologies

To better understand the capabilities and limitations of systems like that shown in Fig. 2, we have been exploring different silicon photonics circuit topologies. A fundamental underlying principle was to minimize per-bit energy costs of the optical communication at a given performance target, which reflects real-world energy constraints in data centers.

In modern high-speed serial links, per-bit energy costs are often dominated by generating high-frequency, appropriately aligned transmit and receive clocks. Macrochip interconnects reduce this overhead in two ways. First, the macrochip—physically bounded by a 12" silicon waveguide lattice—can leverage a characteristic of small systems, that all chips can share the same reference clock from a common crystal. A macrochip is a *mesochronous* system, with bounded phase error between any two chips' clocks. Receivers on an optical link need only adjust phase, a simpler—and mildly less power-hungry—operation than the full clock recovery required of large systems that stretch across boards and racks. Second, and more important, is an argument of physical density: this phase adjustment (and transmit clock generation) is simpler than in traditional serial links, because a macrochip can afford to run links at the CPU clock rate, or double that. This is because WDM, with fine-pitch optical proximity communication, significantly lowers the incremental cost for each added link, obviating the need to overclock a small number of links. Instead, a macrochip can use many slower links for the same bandwidth.

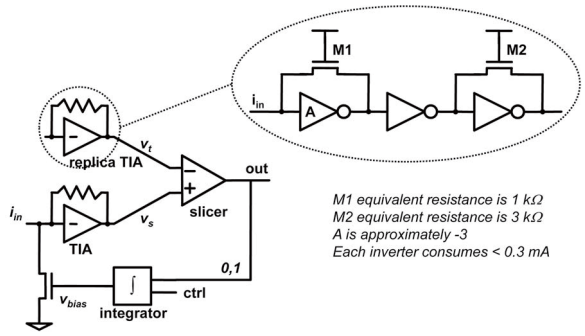

Figure 3: TIA with digital calibration path. The TIA uses n-transistors in triode as feedback resistors. The integrator consists of an adder, two 8b registers storing calibration voltages, and an output DAC.

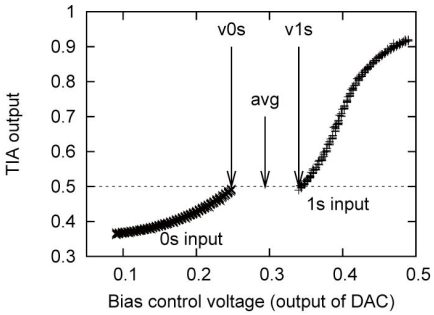

Figure 4: Calibration first finds v0s, the bias that forces a 0 input to the threshold voltage v_t. It then finds v1s, the bias that drives a 1 input to the same v_t. The final bias is the average of v0s and v1s.

Macrochip optical links are single-ended to conserve energy; however, single-ended links often require DC balanced data, with an equal number of 1s and 0s. This allows using the long-term average of input bits as a binary decision threshold in a voltage slicer. Using a DC balanced code like 8B10B is costly: it imposes not only encoding/decoding latency but also a 25% bandwidth (and therefore energy) overhead, and requires generating and distributing a second clock, 25% faster than the chip clock. The macrochip dispenses with DC balanced data in favor of a periodic refresh scheme, detailed below, during which all receivers can recalibrate to proper input levels. The small size of a macrochip helps again, by making it plausible to distribute and synchronize on a global "refresh" signal. This refresh process can be held to well under a 0.01% bandwidth overhead, although it requires system-level support to periodically suspend global communication.

A. Modulator driver circuits

We designed sub-pJ/bit drivers for both high finesse ring structures and absorption modulators. The former requires a 2 V swing in order to achieve an adequate extinction ratio; the latter may only need a 1 V swing but presents significantly higher capacitive load. For the rings, we built cascoded drivers based on [15] to drive 5 Gbps data on 5-15 μm diameter rings, which present 50-100 fF of capacitive load including bonding parasitics.

B. Receiver circuits

Macrochip links suffer optical losses from proximity couplers, waveguides, muxes, and demuxes in the path. These losses are bounded, but unknown until the system powers on. For this work, receiver specifications were a sensitivity of -15 dBm and a dynamic range of 7 dB for a 5 Gbps optical input with an extinction ratio of 6 dB. Photodetectors considered had responsivities of 0.5-0.8 A/W and bonded loads of 100 fF.

Large photodetector parasitic capacitance and lack of DC-balanced data complicates the use of integrating front-end receivers [16] where photo-current directly charges the capacitance. Voltage headroom limits and concerns about noise performance also led us away from regulated cascode amplifiers [17]. Instead, we use a three-stage transimpedance amplifier (TIA), followed by a sense amplifier, as shown in Fig. 3 [18]. The first amplifier stage is an inverter and not a common-source NMOS amplifier because the PMOS input capacitance is small compared to the photodetector, so adding the PMOS g_m to the gain improves gain-bandwidth product (about 25 GHz for 0.3 mA bias current). This first stage improves the TIA's input pole, while the second and third stages provide voltage gain to overcome offsets in the sense-amplifier. The bias device maintains the TIA operating point and is discussed next. Input referred noise current was calculated to be 1.1 μA, which gives sufficient SNR for a 20 μA input current.

C. Calibration and refresh

The sense-amplifier maintains a nominal threshold voltage of ½ V_{dd} (set by a replica TIA). The receiver must therefore calibrate the bias device to ensure that the average of input 1s and 0s also leads to a TIA output at ½ V_{dd}. Fig. 3 shows a digital feedback scheme to do this, similar to an analog scheme from [18].

Fig. 4 illustrates the calibration process, simulated in HSpice. Upon initiation of global refresh, the bias control voltage is pre-discharged to 0 V (lower left of the plot), minimizing the bias shunt current, and the transmitter sends a constant string of 0s. With no diversion of input current, at our specified sensitivity and extinction ratio, a 0 will be seen as a high input, and because the TIA has an inversion, sliced as a 0. The integrator will gradually step up the bias voltage until the slicer generates a 1, at which point the TIA output will have just risen past ½ V_{dd}. This bias voltage is called v0s in the figure and is stored in an 8b register. A similar process, based on pre-charging the bias voltage to V_{dd}, sending 1s, and gradually lowering the bias voltage, returns that voltage (v1s) that causes the TIA output to fall just past ½ V_{dd}. Taking the mean of v0s and v1s returns a bias voltage that forces the average of a 0 and a 1 to sit at the ½ V_{dd} threshold.

Averaged bias control voltages do not generate averaged bias currents, due to the quadratic i-v relationship of the long-channel bias device. However, this non-linearity cancels, to first order, the non-linearity present in a TIA that uses an NMOS transistor in triode for its feedback resistance.

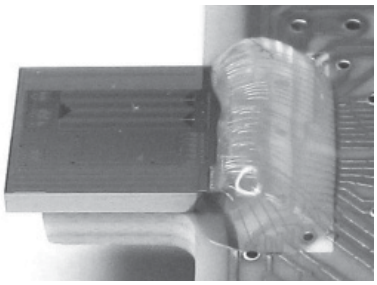

Figure 6: Die photo, showing the chip on a PCB and ready to be face-to-face bonded with any of several optical design of experiments chips. Data and power are delivered at the bottom of the chip and driven up to arrayed transmit and receive experiments.

Note that it is possible that an initial 1 input, with the bias shunt device fully on, is not low enough to be sliced, after the TIA inversion, as a 1. In this case, v1s will be V_{dd}. This represents a very large signal swing and thus a large noise margin in the receiver. The bias shunt device is sized such that the opposite case, with an initial 0 input not high enough, does not occur.

V. PROTOTYPE RESULTS

We built a chip in 90 nm CMOS to test some of these ideas. This testchip does not implement a full optical link; rather, it was designed to be flip-chip bonded (using 15 or 30 μm wide low-capacitance microsolder pads [10]) to one of many optics chips. To characterize wide optical Designs of Experiments, the VLSI chip replicated its modulator driver and receiver circuits in arrayed columns, with some variation in driver transistor sizes. The goal of the chip was to provide a broadly useful VLSI interface, able to meet the worst-case optical device specifications at low energy and area costs, and to be bondable to many generations of optical test devices.

A photo of the chip, wirebonded to a daughter card, is shown in Fig. 6. One of several optical test devices will be attached face down to this chip; results from these tests are being collected now and will be the focus of a future publication. Results on this chip were collected through direct face probing of output driver bond pads and input receiver photodetector bond pads; these helped assure that the chip could be a generic driver or receiver for many different optical device configurations.

Modulators were tested by feeding 5 Gbps PRBS (2^{31}-1) data into the chip using high-speed input pads. Model 35 Picoprobes pulled the signal off of optics bond pads; the probe loading (50 fF) closely modeled the intended ring modulator's bonded capacitive loading. A design error on clock loading led to deterministic jitter, which can be seen in the eye diagram of Fig. 7 as multiple crossings. No errors were seen in over 10^{13} bits sent. Power measurements were limited by multimeter precision, but total power was bounded by 3 mW.

Receivers were tested at DC to check the refresh mechanism, and at speed by using a Model 10 Picoprobe to force a voltage into the TIA input and reading data off the high-speed output pads. No errors were seen in > 10^{12} bits sent. Power (static plus dynamic) was measured to be under 2 mW, not including clock phase adjustment.

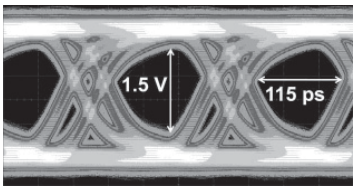

Figure 7: Probed transmitter data eye. Clock skew leads to data-dependent deterministic jitter. Eye shows 2 V modulator drive at 5 Gbps, measured BER was under 10^{-13} and total power under 3 mW.

REFERENCES

[1] A. Narasimha *et al.*, "A fully integrated 4x10-Gb/s DWDM optoelectronic transceiver implemented in a standard 0.13 μm CMOS SOI technology," IEEE J. Solid-State Circuits, vol. 42, no. 12, pp. 2736-2744, December 2007.

[2] M. Asghari, "Silicon photonics: a low cost integration platform for datacom and telecom applications," Conference on Optical Fiber Communications, OFC/NFOEC, pp. 1-10, February 2008.

[3] Y.-H. Kuo *et al.*, "Strong quantum-confined Stark effect in germanium quantum-well structures on silicon," Nature 437, no. 7036, pp. 1334-1336, October 2005.

[4] C. Batten *et al.*, "Building manycore processor-to-DRAM networks with monolithic silicon photonics," IEEE Symp. Hot Interconnects, pp. 21-30, August 2008.

[5] M. Lipson, "Guiding, modulating, and emitting light on silicon—challenges and opportunities," IEEE J. Lightwave Technologies, vol. 23, no. 12, pp. 4222-4238, December 2005.

[6] R. Ho *et al.*, "The future of wires," Proc. IEEE, vol. 89, no. 4, pp. 490-504, April 2001.

[7] R. Ho, *et al.*, "High speed and low energy capacitively driven on-chip wires," IEEE J. Solid-State Circuits, vol. 43, Jan. 2008.

[8] B. Kim *et al.*, "A 4 Gb/s/ch 356 fJ/bit 10mm equalized on-chip interconnect with nonlinear charge-injecting transmit filter and transimpedance receiver in 90nm CMOS," IEEE Int'l Solid-State Circuits Conference, pp. 66-67, February 2008.

[9] R.T. Chang *et al.*, "Near speed of light signaling over on-chip electrical interconnects," IEEE J. Solid-State Circuits, vol. 38, no. 5, pp. 834-838, May 2003.

[10] A.V. Krishnamoorthy *et al.*, "Computer systems based on silicon photonic interconnects," Proc. IEEE, vol. 97, no. 9, July 2009.

[11] R. Drost *et al.*, "Proximity Communication," IEEE J. Solid-State Circuits, vol. 39, no. 9, p. 1529-1536, September 2004.

[12] J. Mitchell *et al.*, "Integrating novel packaging technologies for large-scale computer systems," ASME InterPACK, July 2009.

[13] I. Shubin *et al.*, "Novel packaging with rematable spring interconnect chips for MCM," Electronic Components and Technology Conference, May 2009.

[14] J. Cunningham, *et al.*, "Optical proximity communication in packaged Si photonics," IEEE Conf. Group IV Photonics, Sept. 2008.

[15] S. Palermo *et al.*, "High-speed transmitters in 90nm CMOS for high-density optical interconnects," ESSCIRC, pp. 508-511, September 2006.

[16] A. Emami-Neyestanak *et al.*, "A 1.6 Gbps, 3 mW CMOS receiver for optical communication," IEEE VLSI Symp., pp. 84-87, June 2002.

[17] S.-M. Park *et al.*, "1.25 Gbps regulated cascode CMOS transimpedance amplifier for gigabit Ethernet applications," IEEE J. Solid-State Circuits, vol. 39, no. 1, pp. 112-21, Jan 2004.

[18] M. Ingels *et al.*, "A 1 Gbps, 0.7 μm CMOS optical receiver with full rail-to-rail output swing," IEEE J. Solid-State Circuits, vol. 34, no. 7, pp. 971-977, July 1999.