INVITED PAPER

Device Requirements for Optical Interconnects to Silicon Chips

Optics may allow interconnects to continue to scale to match the processing ability of future electronic chips, though very-low-energy optoelectronic devices and novel compact optics will be needed.

By David A. B. Miller, *Fellow IEEE*

ABSTRACT | We examine the current performance and future demands of interconnects to and on silicon chips. We compare electrical and optical interconnects and project the requirements for optoelectronic and optical devices if optics is to solve the major problems of interconnects for future high-performance silicon chips. Optics has potential benefits in interconnect density, energy, and timing. The necessity of low interconnect energy imposes low limits especially on the energy of the optical output devices, with a ~10 fJ/bit device energy target emerging. Some optical modulators and radical laser approaches may meet this requirement. Low (e.g., a few femtofarads or less) photodetector capacitance is important. Very compact wavelength splitters are essential for connecting the information to fibers. Dense waveguides are necessary on-chip or on boards for guided wave optical approaches, especially if very high clock rates or dense wavelength-division multiplexing (WDM) is to be avoided. Free-space optics potentially can handle the necessary bandwidths even without fast clocks or WDM. With such technology, however, optics may enable the continued scaling of interconnect capacity required by future chips.

KEYWORDS | International Technology Roadmap for Semiconductors (ITRS) roadmap; optical interconnections; optical modulators

I. INTRODUCTION

Optical fiber has already taken over the task of long-distance communications from electrical cables and is increasingly advancing in connections between different parts of large electronic systems [1]. Substantial recent efforts have focused on using optics on circuit boards [2], but wires still dominate all short-distance communications inside information-processing machines, especially on integrated circuit chips and on circuit boards.

As clock speeds and wiring density inside machines have increased, however, interconnection through wires has increasing difficulties [3]–[8]. Now the limited capacity of electrical interconnects is a problem for systems not only on the backplanes and busses between boards but also at the shorter distances between chips and even on chips. Physical arguments for why optics could help interconnections have been presented [9], [10]. Recent systems proposals have advocated optics on-chip through improved architectures enabled by optics [11], [12]. The purpose of this paper is to establish targets for research in optoelectronic and optical devices if optics is to solve the central interconnect problems to and on silicon processing chips.

This paper is unavoidably speculative. We have to try to project where electronic chips and electrical interconnect technology will be in the future and what the demands on interconnects will be. We also have to anticipate the performance of electronic, optoelectronic, and optical technologies that in many cases do not exist yet. Some of the speculation is unavoidably simplistic or even naïve. There is, however, little doubt that interconnects are now and will be increasingly a major limitation on information-processing systems. There is also little doubt that the physics of optics offers potential solutions. At least, we hope to show scaling trends, key technological requirements, and promising opportunities.

In Section II, we start by comparing the underlying physics of electrical and optical interconnects. The energies and densities required in future interconnects are discussed in Section III. In Section IV, we discuss the requirements for optical systems for interconnects, and in Section V we draw some conclusions from these energy and density arguments for optoelectronic devices and systems. We summarize our conclusions in Section VI.

Manuscript received December 2, 2008. First published June 10, 2009; current version published June 12, 2009. This work was supported by the Interconnect Focus Center.
The author is with the Ginzton Laboratory, Stanford University, Stanford, CA 94305-4088 USA (e-mail: dabm@ee.stanford.edu).

Digital Object Identifier: 10.1109/JPROC.2009.2014298

0018-9219/$25.00 ©2009 IEEE

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

II. PHYSICS OF ELECTRICAL AND OPTICAL INTERCONNECTS

We have discussed the comparison between the physics of optical and electrical interconnects previously [6], [9], [10], [13], [14]. Underlying the contrast is the very high carrier frequency of optical signals—on the order of 200–1000 THz for corresponding free-space wavelengths of $\sim$1.5 μm to 300 nm. The short wavelength of light means that dielectric waveguides (which can have very low loss) can be used to guide the waves. Optics therefore avoids the metal waveguides that are essential for confining the radio-frequency waves of electrical interconnects, and hence also avoids the resistive loss physics that dominates the propagation loss and distortion of electrical lines. The very high carrier frequency of optics means that the high-speed modulation of optical beams makes practically no difference to their propagation, at least over the size scale of information-processing machines. Hence, (modulation-) frequency-dependent crosstalk and reflection is avoided; an optical system designed for one signal modulation frequency will work for higher modulation frequencies. Going along with the high frequency and short wavelength is that fact that the photon energy is large ($\sim$0.8 to 4 eV for the 200–1000 THz frequency range), which means that optical signals are created and detected quantum mechanically, in contrast to the classical currents and voltages of electronics. This quantum mechanical nature leads immediately to voltage isolation in all optical interconnects, and to a process called quantum impedance conversion [9], [13] in optical links that could save interconnect power.

These differences in physics lead to three specific major possible practical advantages for optical interconnects.

1) Interconnect Density: Optics avoids a key limit to the density of information that can be sent over relatively long distances. Because of the resistive loss in electrical lines, in lines without repeater amplifiers and for signaling limited by eye closure (rather than, say, the Shannon limit given by noise), the bit rate on electrical lines is limited to

$$B \leq B_o \frac{A}{L^2} \tag{1}$$

where A is the cross-sectional area of the wiring, L is the length of the wires, and B_o is a constant. $B_o \sim 10^{16}$ b/s for the resistive–capacitive lines that are typical on chip, a slightly smaller number for inductive–capacitive lines with resistive loss (RLC lines), and $B_o \sim 10^{17} - 10^{18}$ b/s for off-chip equalized RLC lines presuming that the receiver can operate with up to 20 dB power loss in the line [6]. (The $B_o \sim 10^{17}$ number corresponds to realistic equalized cables or printed circuit board traces and the 10^{18} number is for ideal equalized lines.)

The fact that the ratio A/L^2 is dimensionless means that once we have filled all available space with wiring, the bit-rate capacity of the system cannot be increased by making the whole system either bigger or smaller. To exceed these capacities, some change of interconnect technology is required; electrically we would have to move towards modem techniques to try to approach the Shannon limit to capacity, with a corresponding increase in the electronic complexity and the possibility of increased power dissipation overall.

Since optics does not have this resistive loss physics limiting it, it can be particularly attractive for relatively long lines with high data rates and limited cross-sections. Additionally, because the carrier frequency is so high, there is a very large amount of available spectrum allowing wavelength-division multiplexing (WDM) that could increase the aggregate bit rate of a given optical beam well beyond the modulation rate possible on any one channel.

2) Interconnect Energy: Optics may be able to save energy in interconnection because it is not necessary to charge the line to the operating voltage of the link [9], [13]. When we communicate electrically, we charge up the whole line (or at least a section of it whose length corresponds to the pulse length) to at least the signaling voltage. Though that voltage need not be the logic voltage—low swing signaling is now common for signals on backplanes, for example [15]–[17], and has recently been advocated also for on-chip use [18]–[20]—this can be a significant energy, a total energy of

$$E_s \geq C_l V_r^2 \tag{2}$$

where C_l is the capacitance of the line (or at least the portion charged by the signal pulse) and V_r is the signaling voltage. Since the capacitance of all well-designed electrical lines is similar ($\sim$2 pF/cm or 200 aF/μm) (see, e.g., [6]), this energy cannot easily be reduced other than by reducing voltage swing.

By contrast, optical interconnects use quantum sourcing and detection of the signal, which makes the classical voltage in the medium not directly relevant (quantum impedance conversion [13]). In optics, the relevant energy for comparison instead is the optical energy required to discharge the total capacitance C_d of the photodetector and the electrical input to which it is connected by the required signal voltage, i.e.,

$$E_p \geq C_d V_r \frac{\hbar\omega}{e} \tag{3}$$

where the voltage $\hbar\omega/e$ is numerically equal to the photon energy in electron-volts. (Here for simplicity we assume a photodetector of unit quantum efficiency, i.e., one electron per photon.) The inequality in (3) accounts for

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

the loss in the optical link and the additional energy cost of the optical output device.

Optics has the potential to win here to the extent that $C_d \hbar\omega/e < C_l V_r$. Since the electrical signaling voltage V_r might be small (e.g., 100 mV or less) compared to $\hbar\omega/e$ (e.g., $\sim$1 V), optics wins only if $C_d \ll C_l$ (e.g., by a factor of ten or more). Hence optics can only win in this energy regard if the line is relatively long and the total detector/input capacitance C_d is small. For example, for a hypothetical total input capacitance $C_d \sim 1$ fF in some very tightly integrated photodetector/transistor combination, the line capacitance C_l we are avoiding would need to be at least 10 fF for our example numbers, which corresponds to at least 50 μm length at 200 aF/μm.

This particular 50 μm break-even length for optics is arguably based on very favorable assumptions for optics; not only does it presume a very low detector capacitance but it also neglects the loss in the optical line and the inefficiency of conversion from electrical to optical signals. More detailed studies give longer crossover lengths for the energy benefit of optics [8], [10], [21]–[25] or are more pessimistic [26]–[28]. This argument does, however, show that there is potentially an energy benefit for optics, one that becomes progressively better as we consider longer lines and lower detector capacitance.

3) Clock and Signal Timing: Optics may be able to deliver and retain very precise timing in clocks and signals [29]–[31]. Optical signals, including short (e.g., picosecond) pulses, do not spread substantially in propagating over the size scale of an information-processing machine. Additionally, short pulses can directly deliver very precise timing edges and could have other benefits [32], including reducing latency [33] and improving signal timing [34]. Optics could be useful for reducing the number of levels in the clock distribution tree [24], [30], thereby reducing clock power dissipation and improving jitter, though there is likely not enough available optical power to clock the entire chip [30]. Multichannel (e.g., WDM or parallel free-space array) signals could, however, retain their relative timing, thus avoiding having to compensate separately for timing variations between channels [35]; only one clock channel or one clock recovery would be required for an entire multichannel line—a significant possible benefit for optics. In what follows, we concentrate mostly on energy and density in interconnects, though timing benefits could also be important.

The basic limiting issues of electrical interconnects on chips have been known for some time (e.g., [7]) and led, for example, to the shift to copper wiring on chips to reduce resistance and improve interconnect speed. One technique that is used routinely to avoid some of the limitations on chip is to break the line into smaller segments through the use of repeater amplifiers. Since the length L of any given segment can then be short, the limitation from $B_o A/L^2$ can be avoided, though one price is a low effective signal velocity on such a repeated line (see, e.g., [10]), adding significant signal delay. Electrical designers have equalization circuit approaches to help with limitations of wiring such as signal distortion and loss (see, e.g., [15], [16], [18]–[20], and [36]), but such approaches add complexity and power and cannot ultimately avoid the underlying physics that limits wires. There are also arguments now that optics might particularly enable networks at short distances, for example, for chip-scale multiprocessors, with additional potential energy reduction and performance improvement [11], [12].

Despite these problems of wiring and the arguments in favor of optics for interconnects to or even on the silicon chip, there is essentially no such use today. There are many possible reasons for this absence of short-distance optical interconnects, but certainly cost targets for introduction of optics at short distances are extreme because wires on chips and boards are very inexpensive. Being able to make the necessary optical and optoelectronic components in a low-cost process compatible with silicon electronics may well be essential for any commercial introduction of optical interconnects. Silicon photonics has advanced substantially in recent years, and has demonstrated many of the key components in such integrated processes (see [37]–[39] for recent reviews and collections of work). Still, however, as we discuss below, the requirements on the optical and optoelectronic devices and their integration are very challenging if optics is to be exploited on any large scale at such short distances, and there are some missing pieces in the technology and devices. Below, we attempt to clarify those demands on devices to give clear focus to research efforts to bring about mainstream use of optical interconnects.

III. ENERGY AND DENSITY REQUIREMENTS FOR INTERCONNECTS

A. Power Dissipation in Chips and Information-Processing Systems

Power dissipation in information-processing systems is a major limitation at many levels, including on complementary metal–oxide semiconductor (CMOS) chips themselves. The International Technology Roadmap for Semiconductors (ITRS)[1] states[2] that the amount of heat that can be removed from a chip in a cost-effective manner is about to reach a plateau, saturating at about 200 W, and that power management is now the primary issue across most application segments. The inability to handle higher powers limits the performance of chips.

There are many sources of power consumption in electronic systems. Interconnects are, however, a major and growing contributor. Approximately 50% of microprocessor power was consumed by the interconnect at the $\sim$130 nm

[1]See http://www.itrs.net/Links/2007ITRS/Home2007.htm.
[2]See http://www.itrs.net/Links/2007ITRS/ExecSum2007.pdf.

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

technology node [40] (approximately the technology in the year 2002), and this is expected to rise to ~80%.

Overall power consumption is an issue too in the economics of large systems. The cost of powering a server over its lifetime is now estimated to be comparable to the purchase cost of the server hardware [41]. The power consumption of information technology is now so large that it is starting to be environmentally significant. Data centers alone were estimated to consume on the order of 1% of all electricity in 2005 (1.2% in the United States, 0.8% worldwide) by one estimate [42] and 1.5% of U.S. electricity in 2006 by another [43], with that latter estimated power consumption rising by about a factor of two by 2011 if historical trends continue. The central processing unit (CPU) in one provider's servers consumed between ~27% and ~57% of the total server power in the 2005–2007 timeframe [44]. Presuming 50% of CPU power is in the interconnects, and taking the lower estimate of 27% of the server power in the CPU would still mean that, in the United States, server interconnect power exceeds the total power generated from solar energy in 2007 [45]. The Global eSustainability Initiative (GeSI) estimates [46] that the information and communications technology (ICT) industries in 2002 are responsible for ~0.5 $GtCO_2$ (gigatons of carbon dioxide) emission in a year, out of a total global emission of 40 GtCO2, corresponding to ~1.3% of all emissions. This study estimates the ICT contribution will rise to ~1.4 $GtCO_2$ by 2020, out of a total of 53 $GtCO_2$, that is, 2.6% of carbon emissions. This report advocates that ICT will lead to substantial overall savings in carbon emissions because of the efficiencies ICT will enable in other areas, but still the overall emissions associated with ICT are significant in their own right.

Power dissipation, including a substantial contribution from interconnects, is therefore a problem that directly limits the performance of chips and increasingly is a significant factor in system economics and the environmental impact of information technology. Arguably, then, it would be very difficult to introduce a new solution for interconnects (such as optics) if it takes more power than the existing (electrical) approach, even if it promises other advantages.

B. Energies and Interconnect Densities for Interconnects to and on Chips

To understand the targets for optical interconnects, we need to understand the energy dissipations and densities of electrical interconnects. (See also [8] and [21]–[23] for other analyses of the relative benefits of optics and electronics in energy and interconnect density.)

1) Off-Chip Interconnects on Boards and Backplanes—Electrical Interconnects and Current Systems

a) Current performance: First, let us look at current technology and demonstrated performance for off-chip electrical interconnects to backplanes or chip-to-chip connections on boards. Several authors summarize recent results on transceiver energies for high-speed (i.e., ~4 Gb/s or faster) off-chip electrical interconnects [17], [47], [48]. Typical results have energies per bit of 2–30 pJ/bit in recent demonstrations. (1 pJ/bit is the same as 1 mW/(Gb/s); the latter is a more common way of stating the unit in the electrical interconnect literature, though the former relates more obviously to the physics of the interconnect devices.) The best current results for transceivers are ~2.8–6.5 pJ/bit for board or backplane interconnects [17] and ~2 pJ/bit [16] for moderate length chip-to-chip interconnects with a relatively ideal electrical channel. Other recent work shows receivers for such links with ~1 pJ/bit at ~10 Gb/s rates [49], [50]. Capacitively coupled proximity communication directly between chips allows particularly high densities of interconnections with similarly low energies [51], and there is a variety of other approaches also for dense short vertical "three-dimensional" (3-D) connections between chips or active circuit layers [52]. Recent work [53] shows energies as low as 80 fJ/bit in such 3-D capacitive connections in 130 nm silicon technology, for example, for face-to-face chips.

We have argued above that it is unlikely that any new interconnect technology can be introduced that takes more energy than the existing electrical approaches. The above electrical energy-per-bit numbers alone strongly suggest that if optical technologies are to take over a substantial fraction of off-chip interconnects on boards or backplanes, then the total (on-chip) system energy to run the optical interconnect cannot exceed ~1 pJ/bit. (See also the recent discussion of "energy per useful bit"—a metric that also factors in interconnect delay—by Krishnamoorthy *et al.* [55], which also advocates a 1 pJ target for this related metric.) Just as in electrical interconnects, there are many energy contributions other than the output device or line driver, so the energy per bit for any optical output device (modulators or light emitters) should be $\ll$1 pJ. To be sufficiently competitive to justify the introduction of optical interconnects, arguably we should require a reduction in overall energy consumption with the optics. Given that it may still be some time before introduction of optical interconnects to the chip as a mainstream interconnect technology, and at least the circuits used for electrical interconnects will continue to improve with the continuing improvement in silicon CMOS technology, arguably we should target system energies in the range of ~100 fJ/bit for optics. Such a system energy per bit argues for optical output device energies in the scale of 10 fJ/bit to a few tens of femtojoules per bit.

b) Future targets: To understand future energy targets for off-chip interconnects, we can look first at numbers from the ITRS roadmap. In Figs. 1–4, we graph some key numbers from this roadmap together with some other results we calculate below, and we summarize some key numbers in Table 1 for two specific target years, 2015 and 2022, as representative examples. For the later

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

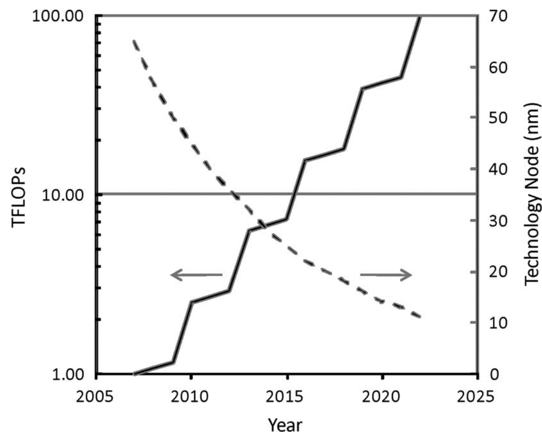

Fig. 1. ***The technology node (a characteristic feature size) for silicon CMOS, and the projected number of FLOPs (units TFLOPs = 10^{12} FLOPs), scaling by the product of the number of transistors times the on-chip clock rate , from a presumed 1 TFLOP in 2007 ([54]).***

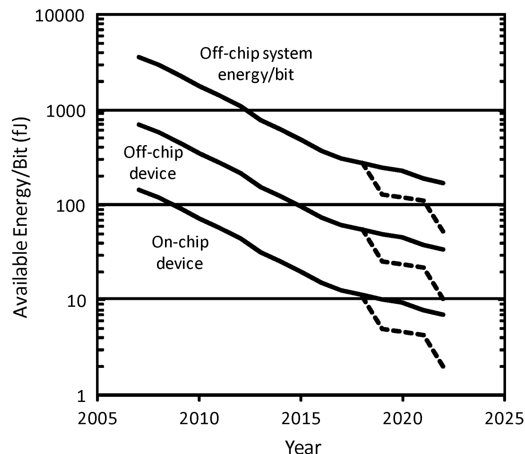

Fig. 3. ***The available energies per bit for interconnects, including the total available system energy per bit for off-chip interconnects (top lines), the energy per bit available for the optical output devices to drive the off-chip interconnect (middle lines), presumed to be 20% of the system energy per bit, and the energy per bit available for optical output devices to drive the on-chip global interconnects (bottom lines). The global on-chip interconnects are presumed to have five times the off-chip bandwidth. The solid lines presume the bandwidths from the product of the off-chip clock rate and the number of signal pins from the ITRS roadmap . The dashed lines presume that the number of bytes of off-chip interconnect per floating-point operation (i.e., the number of bytes/FLOP) is to be maintained in the later years. The middle lines also happen to represent the system energy/bit for on-chip global interconnects because we take that energy also to be five times the device energy for on-chip global interconnects.***

years, we also add projections assuming we retain constant numbers of bytes of off-chip interconnect per floating-point operation (bytes/FLOP or B/FLOP). All the numbers except the energies per bit, the floating-point operations per second (FLOPs), and the numbers at later years for constant bytes per FLOP come directly from the ITRS roadmap , and we use the numbers for high-performance application-specific integrated circuits (ASICs) from that roadmap. The ITRS numbers are based on some presumed scaling of technology or demand, though it is by no means clear that any evolutionary electrical approach would enable these numbers, a point made explicitly in the roadmap.

In these figures and in Table 1, for simplicity, we presume 1 bit/s for each hertz of off-chip clock frequency for each pad. [We could argue that 2 bit/s per hertz is achievable with non-return-to-zero signaling, but in compensation we could also argue that differential lines with two lines (and hence two pads per signal) would be required electrically.] Such a number should also be regarded as an

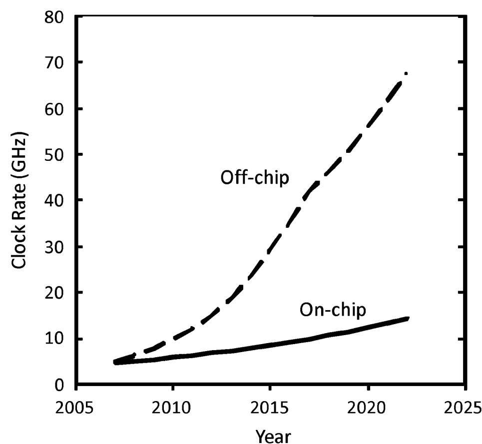

Fig. 2. ***Projected on-chip clock rate and the projected off-chip rate required to drive the chip input and output, according to the ITRS roadmap .***

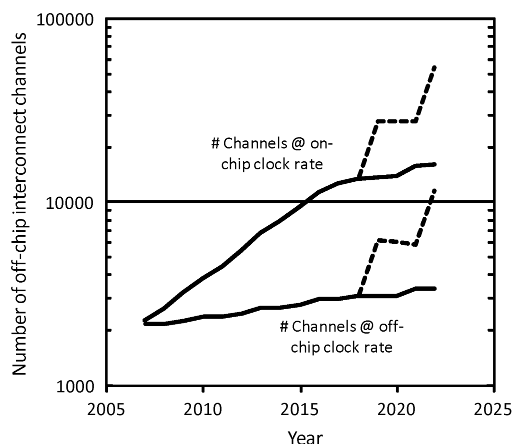

Fig. 4. ***Number of channels required to support the ITRS off-chip interconnect bandwidth [40], presuming either the ITRS off-chip or on-chip clock rate, respectively, for those channels (solid lines). The dashed lines show the corresponding numbers if the number of bytes/FLOP is maintained in the later years.***

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

Table 1 Key Parameters for Technology Node, On- and Off-Chip Clock Frequencies, and Signal Pins From the ITRS 2007 Roadmap, and Calculated Total Input/Output (I/O) Data Rates and Available Energy per Bit for Off-Chip Interconnects, Assuming 20% of Chip Power Is Used for These

Year	Tech. Node (nm)	On-Chip Clock (GHz)	Off-Chip Clock (GHz)	Signal Pins	Total I/O (Tb/s)	fJ/bit for Off-Chip
2007	65	4.7	4.88	2200	11	3500
2015	25	8.5	29.1	2800	82	490
2022	11	14.3	67.5	3420	230	170
2022 constant B/FLOP	11	14.3	-	-	780	50

upper bound, since it is based on the naïve assumption that all the signal pins could be handling high-speed I/O and all could be running at the maximum rate. We should note too that on the ITRS roadmap, achieving the off-chip clock rate for the later years is considered a problem for which manufacturable solutions are not known. Such an increased clock rate is projected in part because the number of pins on a chip is projected to grow only slowly (see [1] for a discussion of the projected relative growths). Nonetheless, this simple product of off-chip clock frequency and number of pads can be useful for scaling arguments.

To calculate the available system energy per bit for the interconnect in these figures and in Table 1, we make the arbitrary assumption that 20% of the total chip power is allocated to the off-chip interconnections, and another 20% to the on-chip interconnect. Given that we also need power for clocking and the logic operations themselves, these are arguably reasonable fractions. It is certainly difficult to argue that they should be substantially higher.

An alternative projection of off-chip interconnect requirements is to ask the capacity of that interconnect to keep up with the ability of the chip to perform computational operations. Such a projection would therefore be attempting to retain a given number of bytes/FLOP. The number of bytes (of communication to memory) per FLOP is a common metric in computer architectures [56], [57], with 1 byte/FLOP being a desirable number (though an increasingly difficult one to achieve) for connections to large amounts of memory. For example, Drost *et al.* [57] summarize bandwidths to different levels of the memory hierarchy in large machines. Those machines surveyed have ~1–10 bytes/FLOP for the connections to local cache memory, falling to ~0.02–0.5 bytes/FLOP for connections to the large, more distant memory. A related idea of the constancy of the ratio between processor power (in instructions per second) and I/O bandwidth (in bits per second) at 1 bit of I/O per instruction—known as Amdahl's balanced system law [59], [60]—is common in discussions of computer design [59], [60].

One recent experimental 275 mm^2 multiprocessor chip [54] performed 1 TFLOP operations per second, in 65 nm node technology, with a 4.27 GHz clock. Such a chip is broadly comparable to a hypothetical "ITRS 2007" 310 mm^2, 4.7 GHz clock-rate ASIC chip on the ITRS roadmap for the year 2007, which would have 2200 off-chip signal pads running at 4.88 GHz, corresponding to an upper bound of 11 Tb/s, or 1.3 Tbyte/s. Hence, dividing 1.3 Tbytes/s by 1 TFLOP, such a chip could hypothetically achieve ~1 byte/FLOP on our simplistic estimates of off-chip bandwidth from ITRS numbers. (Here, for simplicity, we consider one byte of communication as being either one byte going on chip or one byte coming on, i.e., our 1.3 Tbyte/s here is the sum of the rates on and off the chip.)

We can obtain a simple estimate of the capability of future chips to perform floating point operations using the ITRS predictions. The ITRS 2007 chip had an estimated 1106 million transistors per chip in 2007, scaling up by a factor of four to 2015. If we also scale the on-chip clock from 4.7 to 8.522 GHz, as suggested by ITRS—a factor of 1.8—then we predict a chip with $\sim 4 \times 1.8 \times 1$ TFLOPs = 7.2 TFLOPs in a simple scaling in 2015. The ITRS projections (as calculated above) give a presumed ~11 Tbyte/s bandwidth in 2015, so the ability to provide ~1 byte/FLOP (actually 11/7.2 bytes/FLOP) would be retained for this hypothetical chip. Scaling to 2022, ITRS has the number of chip transistors increasing by a further factor of eight (to 35 391 million), and the on-chip clock rate increasing to 14.343 GHz, a further factor of 1.68. The chip performance on a simple scaling would increase to $8 \times 1.68 \times$ 7.2 TFLOPs = 96.8 TFLOPs. We have plotted these simple scaling projections in Fig. 1. The ITRS projections, however, have an upper bound of only ~29 Tbyte/s off-chip bandwidth, corresponding to only ~0.3 byte/FLOP, so the architecture would be significantly impacted by the lack of interconnect bandwidth. We see this discrepancy in the contrast between the solid and dashed lines in Fig. 4. To achieve 1 byte/FLOP would require ~780 Tb/s of off-chip bandwidth.

Whether or not current or future chips can in practice achieve a number as high as 1 byte/FLOP in connecting to off-chip memory is speculative. It is clear, though, that the off-chip interconnect in ITRS projections does not keep up with the ability of the chip to perform logic operations.

Another key point about such off-chip interconnect bandwidths is the energy available per bit. The ITRS roadmap has the power dissipation of chips saturating at ~200 W. Then for the off-chip bandwidth of 82 Tb/s of the ITRS 2015 chip, we have only ~490 fJ per bit of available energy; and, for the 230 Tb/s of the ITRS 2022 chip, we have only ~170 fJ. To have an interconnect that would keep a constant ratio of bytes/FLOP for the ITRS 2022 chip, we would have only ~50 fJ/bit available. These energies are the total system energies available per bit. For a hypothetical optical interconnect, we can only allocate a portion of that to the optical output device. In Fig. 3, we have plotted calculated off-chip optical output device energies assuming that the device consumes 20% of the

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

system energy per bit. On this basis alone, that energy is 97 fJ by 2015, falling to 34 fJ by 2022. To retain the ratio of bytes/FLOP, that energy falls to 10 fJ by 2022.

Given the current state of the art in electrical off-chip interconnects, which is in the range of 2 pJ or greater, arguably such electrical approaches with future technology might achieve the required ~490 fJ/bit in 2015. Whether electronic technology could handle further reductions in off-chip energy is a more open question. Such questions are the subject of ongoing research in electrical interconnects.

A second question for electrical interconnects is whether the off-chip wiring would have enough cross-sectional area to handle the signals. One upper bound guideline for electrical interconnect densities is (1), which can project the minimum cross-sectional areas for connections of a given length when performance is limited solely by loss and distortion in high-quality lines and the lines are equalized with appropriate electronic circuits. Equation (1) does not, therefore, account for the additional problems of reflections and distortions from junctions, bends, and interlayer via connections in boards. It presumes point-to-point connections (i.e., no bus structures with multiple taps) and presumes the low-frequency skin effect with bulk copper conductivities. It also neglects dielectric loss. With a value of $B_o \sim 10^{17}$ b/s, this bound would project that the cross-sectional area of wiring required might just be possible for 10 cm lines on boards for the 2015 and 2022 ITRS chips, at least if there are only a few chips to be connected in this way. Whether the backplane could provide enough cross-sectional area for 1-m-long wiring is more doubtful. If we were to ask for the 780 Tb/s of the hypothetical 1 byte/FLOP chip of 2022, for example, for 1-m-long connections the wiring would need a cross-section of at least ~ 80 cm^2, which appears quite unrealistic. The issue of reducing the cross-sectional area of wiring is already one of the major reasons pushing the implementation of optical interconnects in larger systems [1], and we can expect this will continue to be a major practical reason at shorter distances also.

2) On-Chip Interconnects—Electrical Interconnects and Current Systems: Electrical wires have been, and will continue to be, extremely convenient and effective for on-chip interconnections. Their manufacturing cost is very low, and extremely complex multilayer interconnections can be fabricated.

The operating energies are low for all short interconnects; at ~2 pF/cm capacitance, even at 1 V signaling, the CV^2 energy cost is only ~200 aJ per bit for each micrometer of line length, which corresponds to ~20 fJ per bit for a 100-μm-long line. For longer lines, however, the energies, at least for such simple "on/off" full-swing signaling, do start to become significant on the levels of energies we have been discussing. Simple 1 V signaling across a 2 cm chip would cost ~2 pJ per bit just for charging and discharging the line.

Density of interconnects is not a problem for short lines. For longer lines, breaking the line up into small segments with repeater amplifiers can avoid the bit-rate density limits such as those given by (1) (see, e.g., [3] and [10]), though this can lead to low effective signal propagation velocities and hence significant delays [3], [10]. Such repeater amplifiers also do not reduce the energy to send a bit. One key question in on-chip interconnects is whether energy per bit can be reduced for the longer "global" interconnects while still retaining sufficient density and limiting the delay in the interconnect lines.

The energies of on-chip electrical interconnects for the "global" lines with lengths in the range of 5 mm or more could be reduced from the ~ picojoule numbers calculated above through the use of low-voltage signaling. For example, recent work on on-chip interconnects has simulated a 1-cm-long complete interconnect, including clocking, at ~1 pJ per bit [18] in 90 nm silicon technology, using low-voltage differential signaling.

Another recent interesting suggestion to reduce on-chip electrical communication energy per bit is to use equalized lines on chip [19], [20]. Kim and Stojanovic [19], [20] have analyzed optimized on-chip lines, both with repeaters and with equalization circuits, for 5–15 mm lengths in 90 [19] and 32 nm [20] (year 2013) technology. The energy per bit for such a system depends greatly on the bandwidth density. Kim and Stojanovic consider densities of bits/second per micrometer of wiring layer width in a given wiring layer. Their numbers can be converted to bits/second per unit wiring layer cross-sectional area by dividing by the wiring layer total thickness. For example, in metal interconnect level 9 (M9), the total thickness of the wiring layer, including dielectrics, is ~1.4 μm. They project, for example, that 15-mm-long on-chip repeated lines will consume ~300 fJ per bit in 32 nm technology for densities up to $\sim 1.5 (\text{Gb/s})/\mu\text{m} (\equiv 1.1\ (\text{Gb/s})/\mu\text{m}^2)$, with larger energies for lower metal levels. At the same bit/second densities, they project that equalized lines would consume ~220 fJ/bit, though that energy would drop to ~54 fJ/bit for $\sim 0.5\ (\text{Gb/s})/\mu\text{m} (\equiv 0.35\ (\text{Gb/s})/\mu\text{m}^2)$ bit-rate density. Increasing bit-rate densities beyond these numbers would apparently lead to substantial growth in energy for these equalized systems, so these bit-rate densities might be considered approximately the largest ones that lead to significantly lower energies compared to repeated lines.

We can compare these results to an equation of the form of (1). If we put a bit-rate density of 1.1 $(\text{Gb/s})/\mu\text{m}^2$ for a 15-mm-long line into (1), then we obtain $B_o \cong 2.5 \times 10^{17}$ b/s. Interestingly, this is a comparable number to the B_o values calculated for off-chip equalized RLC lines (for which $B_o \sim 10^{17} - 10^{18}$ b/s), so this formula may be a useful empirical guideline also for these on-chip lines.

If we presumed we occupied the entire M9 layer of a 20 × 20 mm chip with interconnects running at this density of 1.5 $(\text{Gb/s})/\mu\text{m}$, then the total bit rate running through those interconnects would be ~30 Tb/s. For

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

shorter lines, Kim and Stojanovic estimate larger possible bit-rate densities and smaller energies—e.g., 3.5 (Gb/s)/μm with 130 fJ per bit for 10 mm lines in M9. Whether these are sufficient numbers for a high-performance chip at the 32 nm technology node of year 2013 is an open question. The off-chip aggregate data rate projected for the ITRS 32 nm node in high-performance ASICs is 2808 data pins running at 18.63 GHz = 52 Tb/s, which is broadly comparable to our calculated 30 Tb/s on-chip data rate for long wires for that node, and so such equalized electrical interconnect might provide the on-chip densities at least to drive the off-chip bandwidth for that node.

To answer the question of whether such electrical on-chip interconnects could continue to handle the necessary on-chip capacity, we would need to answer the architectural question of how much on-chip data over what distance is needed for future high-performance chips. Wire length analysis has been performed for single microprocessors (see, e.g., the ITRS roadmap[1]). This analysis may, however, be of limited use for our purposes because it appears that on-chip architectures may be changing to ones with networks of large numbers of processors [chip multiprocessor (CMP) architectures]. As this happens, the individual processors become physically smaller with future generations, and so the physical wire length in them shrinks. If the clock rate in those processors remains substantially constant as they are shrunk—a concept that limits power dissipation—then the electrical interconnects within each processor can continue to work at the same bit rates as before while becoming shorter physically and hence consuming less energy per bit. If and when that architectural change takes place, the longer interconnects become those in the network that connect the individual processors, and so we should look at those networks to get another perspective on the issues for the longer wires on the chip.

As a representative example of future on-chip interconnect requirements for longer links, consider the hypothetical CMP of Owens *et al.* [60]. This chip, based on 2015 technology, at the 22 nm node, presumes a 16 $\times$ 16 grid of 256 processor cores on a 400 mm^2 die. A mesh routing network with a total of 480 links each running at 1 Tb/s aggregate data rate connects each core to its neighbors. Each link consists of 144 physical wires, each 1.25 mm long, and each running at a chip clock rate of 7 GHz. The desired total power to run this network is 10 W. With a total bit rate of 480 Tb/s on all the links, there is, therefore, only 10 W/480 Tb/s = 21 fJ/bit available on this design. Their estimate of the electrical power to run such links, based on wires with repeater amplifiers, is 312.5 fJ/bit. Hence immediately we see that electrical power dissipation in the networks in CMP architectures is likely to be a substantial problem for aggressive future chip designs. Running the links just for this mesh network at the 312.5 fJ/bit would consume the entire 150 W chip power budget.

Note, incidentally, that this total on-chip bandwidth of 480 Tb/s for these moderately long (1.25 mm) links is about six times larger than the 82 Tb/s off-chip bandwidth capacity calculated above for the ITRS 2015 chip, though the bisection bandwidth (the bandwidth crossing a line across the chip) is targeted at 16 Tb/s for this example CMP chip [60].

Just what the relation should be between off-chip bandwidth and global on-chip bandwidth is not clear without making additional architectural and application assumptions, though it would be difficult to imagine that the on-chip global bandwidth could be significantly less than the off-chip bandwidth—the data going in and out of the chip has to go to and come from somewhere on the chip, and presumably those destinations and sources are not just around the edge of the chip. The numbers quoted above from Drost *et al.* [57], with ~1–10 bytes/FLOP for connections to local cache memory suggest on-chip bandwidths are likely much higher than off-chip bandwidths, perhaps as much as a factor of ten.

Hence, we should expect on-chip global bandwidth to be at least comparable to the off-chip bandwidth for a given chip, and possibly significantly larger. For illustration, we presume that the on-chip global bandwidth is five times larger than the off-chip bandwidth; though this factor is somewhat arbitrary, it is consistent with the 1–10 bytes/FLOP from Drost *et al.* [57], and with the 480 Tb/s on-chip bandwidth of Owens *et al.* [60] for a chip we calculate to be capable of ~82 Tb/s off-chip interconnect. Hence, we take the targets for energies per bit for on-chip global interconnects to be ~5 times smaller than those for off-chip interconnects. Consequently, where we were looking for total system energies per bit ~1 pJ–100 fJ for off-chip interconnects, we could argue that we should be requiring system energies of ~200–20 fJ per bit for global on-chip interconnects. The specific projected numbers for the on-chip system energy/bit from the ITRS scaling based on constant bytes/FLOP are the same as the off-chip device numbers shown in Fig. 3, so those lines also represent the on-chip system energy/bit. This is because we chose factors of 1/5 in both cases (off-chip device energy/bit ~20% of off-chip system energy/bit; off-chip bandwidth 20% of on-chip global bandwidth).

IV. REQUIREMENTS FOR OPTICAL SYSTEMS

Before discussing optoelectronic device requirements in the next section, here we discuss some basic numerical targets required of the optics itself for optical interconnects that would meet the desired bandwidths. There are two broad categories of approaches—so-called "free-space" optics and guided wave optics. WDM is an additional option that may be particularly useful for the guided wave approaches and could be used in free-space systems also [35].

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

A. Off-Chip Interconnects on Boards and Backplanes

Optics can certainly avoid the kinds of density bounds encountered in electrical systems, e.g., as characterized by (1), for off-chip interconnects. One optical fiber, with a diameter of 125 μm, on its own can carry more than 20 Tb/s of information [61], [62] in telecommunications systems, for example.

One key parameter for discussing optical systems is how many physical channels we need to carry the interconnect data rate. This number is the total data rate divided by the relevant clock rate. We plot numbers of channels in Fig. 4, where we show numbers of channels based on both off-chip and on-chip clock rates.

1) Waveguide Approaches: If we intend to take the data on and off the chip using waveguides that are connected around the perimeter of the chip, e.g., as waveguides on the board to which the chip is connected, then we are interested in the number of micrometers of perimeter needed for each of these channels. That number will determine what waveguide sizes are required on the board. For a chip area of 310 mm^2 as presumed on the ITRS roadmap , the total perimeter, assuming a square chip, is 70.4 mm. The resulting available chip perimeter per waveguide is shown in Fig. 5 for various assumptions.

a) Optical fibers: If we were to use conventional optical fibers, which have a diameter of 125 μm, stacked side by side and butted against the chip edge, the calculations of Fig. 5 show that WDM would be essential at all years if we are to meet the full off-chip bandwidths we have calculated from the ITRS roadmap. Without WDM, the width available per channel is less than the fiber diameter. On the other hand, 16-channel WDM together with the ITRS off-chip clock rates in principle would be sufficient for all cases except the 2022 case with constant bytes/FLOP. Thirty-two or 64 channel WDM would give more flexibility in using the perimeter and reduce the number of fibers accordingly. Thirty-two channel WDM with 361 fibers would handle the 780 Tb/s of the 2022 case with constant bytes/FLOP using the ITRS off-chip clock rate of 67.5 GHz. With 100 GHz channel spacing in the WDM, the required spectral width would fit within the telecommunications C-band. Such an approach would allow a direct "chip-to-network" connection.

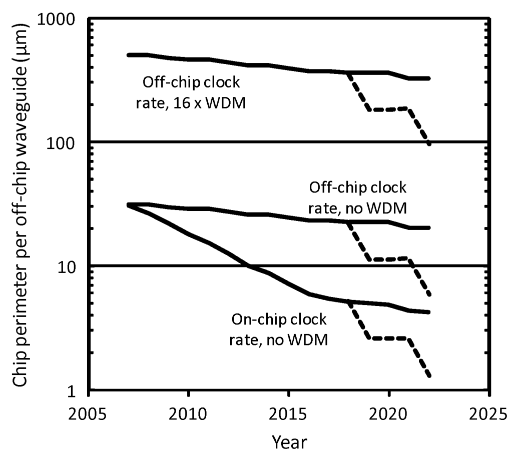

Fig. 5. ***Available width per waveguide channel for waveguides brought off the chip edge. Upper lines—channels running at the off-chip ITRS clock rate, and with 16 WDM channels on each waveguide. Middle lines—channels running at the off-chip ITRS clock rate, with one channel per waveguide. Lower lines—channels running at the on-chip ITRS clock rate, with one channel per waveguide. The solid lines presume the bandwidths from the product of the off-chip clock rate and the number of signal pins from the ITRS roadmap . The dashed lines presume that the number of bytes of off-chip interconnect per floating-point operation (i.e., the number of bytes/FLOP) is to be maintained in the later years.***

There would, however, be substantial device challenges for such a system. Specifically, we would need i) optical output devices (modulators or lasers) capable of running at ~67.5 GHz rates, with low enough optical energies, and ii) compact optical WDMs with 16, 32, or possibly more channels. If we were to position those multiplexers at the edges of the chip, then we would only have ~100–200 μm width for each multiplexer depending on how densely we packed the fibers.

Wavelength splitters are already necessary components in telecommunications WDM systems. See [63] and [64] for discussions of recent integrated technology. The conventional approach for wavelength splitters in guided wave systems is to use arrayed waveguide gratings (AWGs) [65]–[67], but even in miniaturized systems [68], these have centimeter sizes too large for the ~100–200 μm width available here.

Silicon microring resonators allow compact resonators for filters [69] and have demonstrated four-channel WDM operation [70]. Such rings likely have to be individually tuned in practice, however—e.g., using thermal tuning. The power for such tuning would have to be included in the power budget.

An alternative approach would be to find some more compact approach that could split multiple wavelengths at once, perhaps allowing just one or two global tunings of the structure (e.g., center wavelength and overall channel spacing). Recent work in etched echelle gratings [71], [72] is promising for achieving the necessary sizes for such devices. Another promising concept is to use superprism phenomena in photonic nanostructures. In photonic crystal structures, the beam propagation angle can be strongly dependent on wavelength because of group velocity phenomena (see, e.g., [73]–[78]). With photonic crystal structures, the angle can change nonlinearly with wavelength and the beam form can be distorted, however. A more flexible approach is to use custom-designed nonperiodic structures [79]–[83]. These have shown linear dispersion with wavelength, with smaller size than their crystalline counterparts, and can also show controllable (e.g., step-like [81]) forms of the beam shift with wavelength. Recent

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

work [84], [85] has attempted to understand the fundamental limits to how small dispersive optical elements could be made, and the results are promising for future very compact devices.

We certainly would need other optical components for such systems. In particular, we need to be able to couple efficiently from external waveguides, such as fibers, to the chips, and there have been various approaches to such coupling (see, e.g., [64] and [86]–[91]).

Whether or not fibers are used to take the information off each chip, to connect even a small fraction of the information from multiple chips to the world outside the board, optical fibers are likely essential. The use of WDM on those external fibers is likely also essential; otherwise the number of fibers would become too large. Hence, we would need to address the issue of compact wavelength splitters for these external connections even if we did not use them for most of the connections to chips.

b) On-board waveguides: Waveguides in silicon technologies can be made in sizes down to less than 1 μm. References [92]–[96] summarize developments and recent work in waveguides in silicon technology. Hence, we could contemplate chips attached to a larger silicon substrate or "board" containing optical waveguides. Note that the waveguide spacings in Fig. 5 are larger than 1 μm for all cases. Hence, even if we presumed we were taking the waveguide connections off the perimeter of the chip in a single layer of waveguides, we might be able to avoid WDM into and out of those waveguides, and we might be able to modulate at just the on-chip clock rate. The guides would, however, have to be quite small; consequently, loss would become a particularly important parameter.

Of course, any such use of small guides would require very precise alignment between the chip and the silicon optics substrate. An example of a recent approach to such chip to waveguide coupling is given in [89], and various other waveguide coupling approaches [64], [86]–[91] could be applicable here.

2) Free-Space Optical Systems: An alternative to using waveguides to bring the information in and out of the chip is to use "free space" systems that image multiple light beams in and out of the surface of the chips, usually in two-dimensional "surface normal" arrays. Such systems have received substantial research attention and laboratory demonstrations [97]–[104]. In this case, imaging optics is used to relay entire arrays of beams from one chip to another. For example, a six-stage system with over 60 000 light beams was demonstrated using such free-space optics [104]. Some earlier work focused on optical interconnected optoelectronic logic device arrays [97]–[99], and later work has investigated CMOS chips with large arrays of optoelectronic devices hybrid attached to the chips [105], [106]. Much of this work used quantum-well diode structures exploiting the strong quantum-confined Stark effect electroabsorption [107] to make the optoelectronic logic or modulator devices. Device arrays with several thousand elements were demonstrated [106]. Other work used vertical cavity surface-emitting laser (VCSEL) arrays (e.g., [100], [101], and [104]).

Free-space optical systems in general have no basic problem in working with quite large arrays of light beams. Ordinary lens systems can readily handle many millions of resolution elements, including focusing to small spots on the scale of micrometers in size. Combinations of conventional lenses and microlens arrays allow efficient focusing to large but sparse arrays of small spots [103], [108]. Techniques based on planar diffractive optical elements, which are fabricated using lithographic patterning, can generate very large regular arrays of spots from a single laser beam and can implement a variety of quite complex regular interconnection patterns [100], [103], [109]. Though random interconnect patterns are difficult for free-space optics, regular structures such as busses or even strongly interleaved patterns such as perfect shuffles that are suited to operations such as Fourier transforms or switching networks could be quite viable. Microoptical systems can make free-space connections from chip to chip or within a chip [110], and a planar optics technology has been demonstrated that could make entire prealigned "free space" optical systems based on lithography on a planar substrate [109].

Free-space optics has the psychological disadvantage that, in bringing information in and out perpendicular to the surface of the chip, it has a physical architecture quite unlike the planar one of chips, boards, and even optical waveguides. An argument sometimes used against such an approach is that we cannot spare a surface of the chip just for optical beams since we already need one chip surface for wiring and another for heat removal. Perhaps that argument could be resolved by the use of transparent heat conductors or mixing functions on a given side of the chip (such as optical and electrical connections). Silicon substrates themselves are, of course, transparent at the infrared wavelengths used in telecommunications.

Because of the large numbers of light beams that can be handled by free-space optics, even for the most extreme case of the 2022 chip with constant byte/FLOP scaling, there would be no need to use any clock rate beyond the on-chip rate. At the presumed 14.3 GHz on-chip clock rate of 2022 and considering the most extreme case of interconnects capable of constant bytes/FLOP in the later years, approximately 55 000 surface optical "pads" would be required (see Fig. 4). At an example optical pad size of $10 \times 10\ \mu m^2$, the total area consumed by the pads would be 5.5 mm^2, a very small fraction of the 310 mm^2 chip area, so we are not near to any limit of available area. (The pad sizes of such interconnects would be comparable to current 3-D electrical vertical chip-to-chip interconnects [53], but they would not necessarily be restricted to the very close proximity connections of such electrical schemes.)

If we speculate that we could run these surface-normal devices with optical energies ~10 fJ per bit, then the entire

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

interconnect off the chip could be run with one ~8 W laser even for this extreme case. While that is a high power for one laser, it is not inconceivable. The use of a single laser to drive the interconnect also offers the possibility that all the interconnects could be clocked synchronously and the signals retimed by pulsing the laser itself [32], thus possibly saving clock power in the interconnects. Avoiding running any interconnects at the very high proposed off-chip rates of the electrical interconnects (67.5 GHz for the 2022 chip) could also avoid clocking difficulties and additional power dissipation for the time multiplexers and demultiplexers needed for such systems.

B. On-Chip Interconnects

The arguments for optics for on-chip interconnects are less compelling than those for off-chip interconnects, at least if considered on the basis of either energy or bandwidth density alone. Here we will briefly discuss the possible system configurations for optics on-chip and some of the resulting criteria. One benefit optics could bring on-chip is to allow long lines at high bandwidth densities, removing such lines as a physical bottleneck.

1) Waveguides: Waveguides in silicon photonics are one interesting and promising approach [92]–[96]. Even small optical guides on the order of a micrometer in size should have low enough loss that such propagation loss itself is not a substantial issue for chip scales [94]. We see from Fig. 5 that, at least if we take a very simplistic view of waveguides, entirely filling one layer on-chip and connecting from the area of the chip to the edges, possibly we could feed the off-chip interconnect using such a layer of guides with signals running at only the on-chip clock rate and without using WDM in the guides. The resulting guides would become very small and close (a 1.3 μm pitch) in the extreme case of the 2022 chip with constant bytes/FLOP scaling. To give some flexibility, and/or to allow some guides to be used for internal on-chip connections, higher clock rates, multiple layers of waveguides, or some WDM would likely be required.

Whether such guides could handle the internal on-chip interconnects is an open question, which also depends strongly on the architecture. Above, we considered that the on-chip "global" interconnect requirements might be, e.g., five times the off-chip bandwidth, though such connections might be for shorter distances, e.g., 1/5 of the chip size. In that case, possibly one layer of guides would still be sufficient even without higher interconnect clock rates or WDM, just as in the case of guides to feed the off-chip interconnect.

Nanometallic or plasmonic metal waveguides have been considered for on-chip waveguiding in both single-conductor and two-conductor waveguides [111]–[117]. Such waveguides can be very small, possibly even smaller than dielectric waveguides. Such very small waveguides could concentrate light to very small (e.g., < 100 nm scale) device volumes [116], which could be useful, for example, for making very small photodetectors. As waveguides for longer interconnects, while they could be a possible approach, the high loss of small guides means in practice that they would have to be comparable to or larger than dielectric guides to achieve comparable and usable overall loss [111]–[116]. A critical analysis of plasmonic waveguides for interconnect has also been given by Tucker [118]. Hence, while they might have uses, they are not a clearly superior solution for the optical "wiring" at substantially higher on-chip interconnect densities at the scale of the chip.

2) Free Space: Free-space optics could also connect within a chip, though this has received relatively less attention in the research literature. Many of the schemes used between chips could also be used within chips. We saw in the discussion above of chip-to-chip interconnects that the surface-normal optical pads for off-chip interconnects need take up only a small fraction of the surface area, so there is a large amount of area remaining for pads for on-chip interconnects.

Free-space interconnects suit regular interconnection patterns, so it might be particularly interesting for regular on-chip networks, for example, connecting multiple processors on chip. A free-space approach is certainly an interesting option for delivering clock signals synchronously over an entire chip [29]–[34].

V. REQUIREMENTS FOR OPTICAL AND OPTOELECTRONIC DEVICES

A. Energy Targets

We can draw summary conclusions from the energy targets suggested above as follows.

i) To be competitive with the current state of the art in electrical off-chip interconnects, the system energy per bit should be < 1 pJ, and to offer sufficient energy advantage for optics, it should be ~100 fJ/bit or lower.

ii) To meet the demands of off-chip interconnects out to the ITRS projections of 2022, system energies per bit of 100 fJ/bit may sufficient, but to sustain the number of bytes/FLOP in the later years will require 50 fJ/bit or lower system energy.

iii) To be competitive with near-future electrical global on-chip interconnects, the system energy per bit should be $\ll$ 50–200 fJ/bit.

iv) To meet global on-chip interconnect demands out to 2022 will require system energies per bit of ~30 fJ/bit on ITRS projections (assuming the global on-chip bandwidth is 5 times the off-chip bandwidth); and to sustain the number of bytes/FLOP in the later years will require ~10 fJ/bit system energies.

An optical interconnect system needs a transmitter driver circuit, an optical output device (laser or modulator),

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

an optical channel, a photodetector, and a photodetector circuit. Let us consider the photodetector and receiver first.

1) Photodetector and Receiver Circuit: On the numbers we have been discussing here, provided we can make a reasonably efficient and well-integrated photodetector with a low enough capacitance, the photodetector should not pose a basic challenge in meeting these targets. Per square micrometer of area, a 1-μm-thick piece of a typical semiconductor has a capacitance of ~100 aF. Hence photodetectors with areas of a few square micrometers with thickness of ~100 nm or greater will have capacitances of a few femtofarads.

To calculate the total capacitance, we need to estimate the gate capacitance of the transistor with which the detector would be integrated. Calculations based on the publicly available ASU Predictive Technology Model[3] [119], [120] can be used to estimate the gate capacitance of CMOS transistors. For the 90 nm CMOS technology node, the gate capacitance of an NMOS transistor per unit transistor width is estimated to be ~2 fF/μm, and for the 32 nm node the corresponding number is ~1.2 fF/μm, at least for transistors that are wide compared to the gate length (the length in the direction from source to drain). Hence, the gate capacitance per unit gate width decreases somewhat, though slower than linearly, with decreasing transistor size. Because the absolute transistor width used in a given circuit is likely to scale approximately with the gate length, the transistor gate capacitance in a given circuit will tend to scale down substantially, and somewhat faster than linearly with the technology node dimension.

For example, presuming a transistor width of ten times the node dimension in each case (as might be typical in transistors used in analog front-ends so as to minimize the effect of fabrication variations), the NMOS transistor gate capacitance would scale from 1.8 fF for the 90 nm node to 380 aF for the 32 nm node. Hence, provided the transistor(s) and the micrometer-scale photodetector are well integrated, the total capacitance of photodetector and input transistor(s) should be on the scale of a few femtofarads. With only a few femtojoules of received optical energy, at ~1 eV photons in an efficient photodetector (i.e., ~1 electron of current for each incident photon), we would generate a few femtocoulombs of charge, which would swing the photodetector and transistor input by ~1 V, i.e., by a full logic voltage swing. In that case, no voltage amplification would be needed in the front-end receiver circuit—in fact, we could possibly directly drive a CMOS inverter circuit (this is sometimes called a "receiverless" approach [30], meaning there is no receiver voltage amplifier circuit required, and the signals could be fed directly into the logic gates). Since the total energy involved here is a few femtojourles, in a well-designed integrated system, this photodetector/receiver energy need not be a large fraction of the system energies per bit we have discussed. If the optical received energies are somewhat lower or the capacitance is somewhat higher, it is also possible to put in some voltage amplification without greatly increasing the energies [121]. The consequences of somewhat larger detector capacitance have also been considered by, for example, [21]–[23].

[3] http://www.eas.asu.edu/~ptm/.

The idea of such an intimate integration at femtofarad capacitance levels is still slightly speculative. There have been recent demonstrations of very tightly integrated detector/transistor combinations using Ge on Si structures [122]–[124]. Ge is generally an interesting detector material because it has large enough absorption in the near infrared to allow photodetectors with micrometer sizes, and it can be process compatible with silicon. Recent work on Ge detectors includes [125]–[132]. Some of these approaches can use the same structure as is also being used as a modulator, possibly simplifying the fabrication overall [130]. An additional concept for low-capacitance and potentially high-speed photodetectors is to use nanometallic structures, such as antennas [133]–[136] or waveguides [111]–[117] to concentrate light into deeply subwavelength active detector volumes. The first such integration with CMOS technology has recently been demonstrated [136].

2) Optical Output Device and Transmitter Circuit: Perhaps the single largest technological challenge in meeting the energy targets we have suggested lies in the device that converts the electrical signal to the optical one. Historically, all such devices have taken substantial energies to operate, much larger than our targets here. If we can make such a device, the energy to drive it from a well-designed transmitter circuit will be comparable to the device operating energy itself, so we can concentrate on discussing the optical output device energy.

The energy targets we need for these devices can only be a moderate fraction of the total system energy per bit; we also need energy for the transmitter circuit, some energy for the receiver, and energy for other circuit functions, possibly including clocking and clock recovery in the link. There will likely also be other energy losses in the system. Given these other energies required in the total system, we set a target here of the optical output device energy/bit being ~20% of the system energy/bit.

Hence, from the above discussion of systems energy targets, we have optical device energy targets of 10–20 fJ/bit for off-chip interconnects, and ~2–10 fJ/bit for on-chip interconnects. With these targets, we can examine some of the possible device technologies.

There are two general categories—modulators and light emitters. Light emitters themselves could be either incoherent [light-emitting diodes (LEDs)] or coherent (lasers), but we can likely immediately eliminate LEDs. Unless an LED is constrained to emit into only one or a very few spatial modes (as is possible at least in principle

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

through Purcell enhancement in a small, high-Q resonator), it is likely too inefficient optically for coupling into a small photodetector. Uncorrelated light of a given wavelength in different spatial modes cannot be combined back into one mode. Though it is not usually stated in this modal form, this principle is known as the constant radiance (or brightness) theorem, and it is protected by the Second Law of Thermodynamics (if we could achieve this recombination with any passive optical component, we could devise an optical system that would allow us to heat up a hot black body with the combined light from two colder black bodies, thus violating the Second Law). Henceforth, we will consider only lasers or modulators.

a) Directly modulated lasers: A state-of-the-art number for energy per bit in a laser is the 286 fJ/bit demonstrated at 35 Gb/s in a 3-μm-diameter tapered oxide aperture VCSEL [137]. This number is certainly low enough to be quite interesting for present interconnects, even for chip-to-chip use. For future use to meet the targets here for future years, however, this demonstrated number is not low enough. A key question is whether such an approach could be scaled down to the $\sim$10 fJ range of our targets.

For a light emitter, 10 fJ is a very low energy. At 1 V drive, this 10 fJ corresponds to $\sim 10^5$ electrically injected electron-hole pairs. Inverting the population (as is certainly required for lasing) in a single quantum-well layer would typically require $\sim 10^{12}$ carrier pairs/cm^2, or $10^4/\mu$m^2. Hence, the 10^5 electron-hole pairs from our 10 fJ would be just enough to invert $\sim 10\ \mu$m^2 of one quantum well. Given that there will be other energies involved in running such a laser, to achieve operation at 10 fJ total energy levels, the gain volume would have to be significantly smaller than ~ 10 nm $\times$ 10 μm^2 of one quantum well. Hence, likely more aggressive laser structures would be required. One possibility would be photonic crystal resonator lasers with single quantum dot gain regions [138], for example, though such structures are still the subject of basic research. Such nanoresonator lasers can also have potentially very fast modulation speeds [138] despite the usual difficulty with lasers that higher modulation speeds require quadratically higher current densities. (One recent scheme [139] can mitigate such modulation limits with intracavity modulation, however.)

The above lasers are made from III–V materials. In addition to the Group III and Group V materials being dopants for silicon, there is generally the problem that epitaxial growth of III–V compounds on silicon is lattice mismatched, which leads to crystal defects. Such crystal defects notoriously lead to short device lifetimes in forward-biased devices such as laser diodes. One radical possibility is to grow III–V nanowires on silicon [140]; because the wires can be so small, they can avoid the defect formation process.

III–V lasers have been successfully bonded to silicon [141] with evanescent coupling between the laser and silicon waveguiding, though these specific lasers themselves likely still have power dissipation too high to be used in direct modulation for the energies per bit of interest here. Such lasers can also be mode-locked [142] to produce short pulses or possibly frequency combs for WDM use, so they are potentially interesting as optical power supplies.

At the time of writing, there is still no electrically pumped laser in a Group IV material, though there have been reports of gain in silicon nanostructures (see, e.g., [143]), and a combination of tensile strain and heavy doping in Ge may allow net gain [144]. For our interconnect applications, any such laser would also have to be very energy efficient and, if it itself is to be modulated to carry the information, capable of high-speed modulation. As we have shown, the energy targets and modulation speeds are difficult even for the III–V lasers that are the highest performance lasers known.

b) Optical modulators and off-chip lasers: Instead of modulating laser sources directly, we can use modulators as the optical output devices. For modulators, we have two broad categories—refractive and absorptive. To use a modulator, we need another source of light to provide the beam of light they modulate. If that light source is off-chip, we have to make the additional effort to couple that power onto the chip. Putting the source off-chip also has advantages, however. We remove the additional power dissipation of the source itself from the chip. We can centrally control laser wavelength, spectral purity, polarization, and beam form, removing any precise stabilization and control from the relatively harsh environment of the silicon chip. We can use that central laser to clock the entire system [30], [32]. Multiple wavelength "comb" sources are also possible, e.g., by mode-locking [145], giving a set of equally spaced wavelengths suitable for WDM systems. Such a system is also likely easier to fabricate and control if it is off of the chip.

It may be possible to run the entire off-chip interconnect or even possibly the global on-chip interconnect using a single laser, as we mentioned briefly above. Of course, we would also have to deal with coupling losses in such a system.

Refractive modulators: Refractive devices modulate by interfering a beam with itself in some way, either in a single-pass two-beam interference as in a Mach–Zehnder interferometer structure or in some device with multiple interference, such as a resonator. Changing the relative phase of the interfering beams by changing the refractive index changes the output power.

A basic difficulty with refractive modulators is that we have no high-speed mechanism that can usefully give us refractive index changes much larger than $\sim 10^{-3}$. Larger index changes can be induced, e.g., in semiconductors very near to their optical absorption edge (see, e.g., [146]), but then such large changes only occur in the presence of substantial absorption. Even with $\Delta n \sim 10^{-3}$, to get a half-wavelength path-length change at 1.5 μm wavelength

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

would require a device length $L \sim 750\ \mu$m if we changed the index in only one arm. Hence generally Mach–Zehnder modulators have lengths in the hundreds of micrometers or longer.

There has been considerable interest in such two-beam interferometer approaches in silicon photonics (see, e.g., [147]). Because the main refractive index change mechanism available in silicon (free-carrier index change [148]) is relatively weak, such devices necessarily take substantial energy on the scale of interest here. For example, Green *et al.* [149] show 5 pJ/bit at 10 Gb/s in well-optimized 100–200-μm-long devices, much larger than our target numbers.

To make compact refractive modulators, we need to use resonators or possibly slow light [150] to enhance the effect of changing the refractive index in only a smaller length of material. The silicon microring resonator has received much attention (see, e.g., [151]–[156]). Such devices might be able to achieve operating energies in the range of tens of femtojourles and could be very small, e.g., a few micrometers in diameter [152]. The resonators need quite large quality factor (Q) (e.g., $> 10\ 000$ [153]), however, meaning that they have very narrow resonances (e.g., 0.04 nm wavelength range [153]) that have to be precisely tuned. Often that tuning is by temperature, and that temperature would have to be stabilized precisely also to hold the device on resonance (e.g., to a small fraction of a degree Kelvin based on the $\sim 2 \times 10^{-4}$/K temperature dependence of silicon's refractive index [153]). The width of the resonance is also so narrow that it is one of the limits on modulation speed, though modulation above 10 Gb/s is quite possible [151]. An important point in the energy per bit is that the thermal tuning power must be included in estimating the total system power requirement. The required tuning power is not yet clear, but a hypothetical tuning power of 1 mW for a 10 Gb/s modulator would correspond to an additional effective 100 fJ/bit, which would take the energy out of our target range. Such devices do, however, have the advantage that they are automatically also wavelength filters and can perform WDM switching functions as well (see, e.g., [69] and [70]).

Electroabsorption modulators work by changing the optical absorption in a semiconductor structure by applying voltage to it. There are two related mechanisms, the Franz–Keldysh effect, seen in bulk semiconductors and the quantum-confined Stark effect (QCSE) [107], seen in quantum-confined structures such as thin (e.g., $\sim$10 nm) quantum-well layers. These effects are very closely related [155], with the Franz–Keldysh effect being the limit of the QCSE as the quantum-well layers are made thicker. The QCSE has more spectrally abrupt and somewhat stronger changes in absorption coefficient as a consequence of the discretization of the density of states and the stronger excitonic effects in quantum-confined structures.

Both effects require operating electric fields in the range of 1–10 V/μm (10^4–10^5 V/cm). Such fields are readily obtained by reverse biasing $p-i-n$ diodes that contain the bulk semiconductor or quantum well materials in the intrinsic (i) region of the diodes. Both effects are seen near the direct bandgap optical absorption edge and give rise to increases in the optical absorption below the bandgap photon. The Franz–Keldysh effect gives a long, smooth absorption "tail," with typical induced absorption coefficient values in the range of a few hundred cm^{-1}. The QCSE gives more abrupt steps in absorption that shift to lower energy with field, with absorption coefficient values that can be up to several thousand cm^{-1}. With their weaker absorption coefficients, Franz–Keldysh devices are used in waveguide structures that can have the necessary longer interaction lengths. QCSE devices are used in waveguide structures too, but, with their stronger absorption coefficient changes, they can be used for "surface normal" devices of micrometer vertical dimensions (see, e.g., [106]. QCSE modulators are widely used in telecommunications, especially in integrated laser-modulator structures (see, e.g., [156]).

Such electroabsorption modulators do not rely on changing the carrier density in the structure, and so they avoid some of the speed limitations found in directly modulated lasers or in silicon carrier density index shift modulators. They are thought theoretically to have fundamental speed limits well below a picosecond [157]. Quantum-well modulator devices have been tested at high speeds [158], [159] and up to 500 Gb/s effective modulation rate [160].

These electroabsorption mechanisms are very strong; the QCSE may be the strongest high-speed electroabsorption effect, enabling modulators with only a few micrometers of optical path length even without the use of resonators. The performance can also be enhanced by cavities if desired (see, e.g., [161]–[163] for recent examples). Because of the strength of the effects, only modest cavity Q or finesse has been used so far. Devices have also been demonstrated with < 1 V drive swing (e.g., [161] and [163]), as practically required for compatibility with CMOS technology.

Because the mechanisms are so strong, low operating energies are likely possible even without resonators. The core of the operating energy is the energy required to charge up the active volume of the device to the operating field. For a field of 5×10^4 V/cm (5 V/μm), that energy is $\sim$2.5 fJ/μm^3. Even without resonators, devices with energies of tens of femtojoules should be feasible. Some level of temperature stabilization would be required for electroabsorption modulators because the bandgap energy of direct gap semiconductors shifts with temperature (typically ~ -0.4 meV/K) and because any resonators used will also have some temperature dependence from the temperature dependence of the refractive index. Because the electroabsorptive effects can be so strong, however, high Q resonators are not required, and hence these devices are likely much less temperature sensitive than microring resonator refractive devices, for example.

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

Historically, such electroabsorption effects have only been exploited in III–V direct gap semiconductors, though it should be noted that III–V modulators have been successfully grown and operated with good lifetimes on silicon substrates [164]. Reverse-biased structures appear to be much more tolerant of crystal defects. An alternative approach is to bond III–V devices to silicon in a waveguide configuration. For example, Kuo *et al.* [165] have demonstrated waveguide InAlGaAs QCSE modulators bonded to silicon structures.

Recently, interest has turned to electroabsorption in Ge structures grown on silicon. Ge is already known to be process-compatible with silicon CMOS. Though Ge is an indirect gap semiconductor, it does have a strong direct gap near 1.5 μm wavelength. The physics of that direct gap optical absorption is essentially the same as that of common III–V materials. Hence it can show the same electroabsorption effects. The one minor disadvantage is that there is a remaining indirect absorption tail, so such devices have larger loss in their nominally transmitting state than their III–V counterparts.

Strong and clear QCSE was recently observed in Ge quantum wells on silicon, the first time the QCSE was clearly observed in any indirect gap or Group IV material [166]–[168]. The first modulator devices have recently been demonstrated [162], [163], including operation at < 1 V swing [163]. Liu *et al.* [169] have demonstrated a waveguide Franz–Keldysh modulator in a CMOS compatible process, and with an estimated 50 fJ per bit of energy.

Though substantial work remains to be done on optimizing device structures and integration approaches, these Ge devices are very promising for high-speed low-energy optical output devices for optical interconnects to Si. The QCSE devices in particular are promising not only for waveguide devices but also for surface-normal modulators for free-space optical systems.

VI. CONCLUSION

The problems for electrical interconnects to and on chips are significant now and will become very substantial in the future. Optics potentially can address key issues of dense, low-power interconnects and can bring other benefits such as improvements in timing. Here we have specifically addressed the targets for optoelectronic and optical devices if they are to handle the full volume of global on-chip and of off-chip interconnects in high-performance chips and systems. We have required that those optical interconnects have performance that is competitive with or better than electrical interconnects and can scale to future interconnect needs.

The discussion of other issues such as architectures and any detail of the substantial challenges in integration, packaging, alignment, and thermal stabilization and control is also important, but it lies beyond the scope of this paper. We note, though, that optics also has significant potential benefits in clocking and timing interconnects, and optical WDM may also allow novel architectures for on-chip and off-chip networks.

We can summarize the major conclusions of this paper for optoelectronic and optical devices and systems. In considering optoelectronic devices, we have argued first that, just comparing with current and near future electrical interconnect technology, optical interconnects to chips need to target total system energies ~100 fJ/bit to be competitive on energy grounds alone, and that therefore the required optical output device energies should be on the scale of 10 fJ/bit to a few tens of femtojoules/bit. Considering the demands for future interconnects based on ITRS predictions and extensions of those predictions to allow constant scaling of the number of bytes/FLOP leads to similar conclusions on the target energies, with ~10 fJ/bit or less being the required device energy for the later years of the predictions.

The devices would need to work at least at the on-chip clock rate, which scales to 14.3 GHz on the ITRS roadmap; for connections to optical fibers, the higher off-chip clock rates (which scale to 67.5 GHz) would likely be required.

With such energy targets, modulators look to be feasible. Silicon ring resonators might meet the energy targets, though precise tuning of their very sharp resonances and tuning power dissipation are significant issues. Electroabsorption modulators should be able to reach the energy targets, possible even without resonators, though very compact integrated device structures would be required. There also appear to be no basic issues with speed for such electroabsorption devices. Recent developments have opened up substantial opportunities for the strongest electroabsorption mechanisms in Group IV materials on silicon. For lasers to meet the necessary targets as optical output devices, only the most aggressive concepts (e.g., quantum dot nanocavity lasers) appear viable as we look to future interconnects.

Photodetectors would need to be very intimately integrated with their transistors to achieve the most desirable capacitance targets of ~femtofarad or less, though somewhat larger photodetector capacitance may be allowable with the use of receiver voltage amplifiers.

Considering optical systems, a variety of waveguide and free-space approaches could be of interest. Interconnecting the entire bandwidth off the chip with optical fibers around the chip edge appears possible, though high off-chip clock rates and at least moderately dense WDM (e.g., at least 16 channels) would be required. Waveguides on boards might be able to avoid either one of WDM or high clock rates, though they would have to be quite dense for the later years of the projections.

On-chip interconnects might not need WDM and might be able to operate at the on-chip clock rate, though if either higher clock rates or WDM were possible on chip, it would increase the flexibility of the design, and WDM could also enable other network architectures on chip [11], [12].

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

For schemes involving WDM splitters on the chip, the issue of compact WDM splitters remains an open and very important problem. At some point in the system, it will likely be essential to use WDM, at least onto the optical fibers that will carry the information over any longer distances off the boards. Conventional WDM techniques cannot provide sufficiently compact components. Microring resonators are one possible approach, though tuning and tuning power are issues to be resolved. Compact etched echelle grating splitters may be possible. Another radical approach with significant potential is to use superprisms, though this is still very much a research topic.

Free-space optics would have no substantial limits on density or device numbers for interconnection of the surface of the chip, even without WDM or high clock rates, and even out to the most aggressive interconnect requirements of the later projected years. Only some of the optical output devices would be feasible in directly "surface normal" rather than waveguide configurations, though quantum-well electroabsorption modulators have previously been successfully demonstrated for such purposes and may be able to meet energy targets.

Overall, the current understanding of the physics of electrical and optical interconnects, and the many existing and emerging technologies in optoelectronics and optics integrated with silicon CMOS, are very promising for optics to play a substantial role in solving the major looming problems in scaling interconnects for CMOS chips in the coming decades. ■

Acknowledgment

The author is pleased to acknowledge many stimulating and informative discussions with K. Bergman, J. Bowers, L. Coldren, A. Emami, W. Dally, R. Ho, M. Horowitz, J. Kash, P. Kohl, L.-S. Peh, V. Stojanovic, and R. Tucker. He would also like to thank M. Loh and A. Emami for gate capacitance calculations and L. Coldren, A. Emami, and V. Stojanovic for preprints of work in preparation.

REFERENCES

[1] A. F. Benner, M. Ignatowski, J. A. Kash, D. M. Kuchta, and M. B. Ritter, "Exploitation of optical interconnects in future server architectures," *IBM J. Res. Develop.*, vol. 49, no. 4/5, pp. 755–775, Jul./Sep. 2005.

[2] L. Schares, J. A. Kash, F. E. Doany, C. L. Schow, C. Schuster, D. M. Kuchta, P. K. Pepeljugoski, J. M. Trewhella, C. W. Baks, R. A. John, L. Shan, Y. H. Kwark, R. A. Budd, P. Chiniwalla, F. R. Libsch, J. Rosner, C. K. Tsang, C. S. Patel, J. D. Schaub, R. Dangel, F. Horst, B. J. Offrein, D. Kucharski, D. Guckenberger, S. Hegde, H. Nyikal, C.-K. Lin, A. Tandon, G. R. Trott, M. Nystrom, D. P. Bour, M. R. T. Tan, and D. W. Dolfi, "Terabus: Terabit/second-class card-level optical interconnect technologies," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1032–1044, 2006.

[3] R. Ho, K. W. Mai, and M. A. Horowitz, "The future of wires," *Proc. IEEE*, vol. 89, pp. 490–504, 2001.

[4] J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, and J. D. Meindl, "Interconnect limits on gigascale integration (GSI) in the 21st century," *Proc. IEEE*, vol. 89, pp. 305–324, 2001.

[5] J. D. Meindl, "Interconnect opportunities for gigascale integration," *IEEE Micro*, vol. 23, pp. 28–35, May–Jun. 2003.

[6] D. A. B. Miller and H. M. Ozaktas, "Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture," *J. Parallel Distrib. Comput.*, vol. 41, p. 4252, 1997.

[7] K. C. Saraswat and F. Mohammadi, "Effect of scaling of interconnections on the time delay of VLSI circuits," *IEEE Trans. Electron Devices*, vol. ED-29, no. 4, pp. 645–650, 1982.

[8] M. Haurylau, C. Q. Chen, H. Chen, J. D. Zhang, N. A. Nelson, D. H. Albonesi, E. G. Friedman, and P. M. Fauchet, "On-chip optical interconnect roadmap: Challenges and critical directions," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, no. 6, pp. 1699–1705, 2007.

[9] D. A. B. Miller, "Physical reasons for optical interconnection," *Int. J. Optoelectron.*, vol. 11, no. 3, pp. 155–168, 1997.

[10] D. A. B. Miller, "Rationale and challenges for optical interconnects to electronic chips," *Proc. IEEE*, vol. 88, pp. 728–749, 2000.

[11] R. G. Beausoleil, P. J. Kuekes, G. S. Snider, S.-Y. Wang, and R. S. Williams, "Nanoelectronic and nanophotonic interconnect," *Proc. IEEE*, vol. 96, pp. 230–247, 2008.

[12] A. Shacham, K. Bergman, and L. P. Carloni, "Photonic networks-on-chip for future generations of chip multiprocessors," *IEEE Trans. Comput.*, vol. 57, pp. 1246–1260, 2008.

[13] D. A. B. Miller, "Optics for low-energy communication inside digital processors: Quantum detectors, sources, and modulators as efficient impedance converters," *Opt. Lett.*, vol. 14, pp. 146–148, 1989.

[14] D. A. B. Miller, "Optical interconnects to silicon," *IEEE J. Sel. Topics Quantum Electron.*, vol. 6, pp. 1312–1317, 2000.

[15] W. J. Dally and J. Poulton, "Transmitter equalization for 4-Gbps signaling," *IEEE Micro*, vol. 17, pp. 48–56, Jan./Feb. 1997.

[16] J. Poulton, R. Palmer, A. M. Fuller, T. Greer, J. Eyles, W. J. Dally, and M. Horowitz, "A 14-mW 6.25-Gb/s transceiver in 90-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, pp. 2745–2757, 2007.

[17] G. Balamurugan, J. Kennedy, G. Banerjee, J. E. Jaussi, M. Mansuri, F. O'Mahony, B. Casper, and R. Mooney, "A scalable 5–15 Gbps, 14–75 mW low-power I/O transceiver in 65 nm CMOS," *IEEE J. Solid State Circuits*, vol. 43, pp. 1010–1019, 2008.

[18] A. Emami, to be published.

[19] B. Kim and V. Stojanovic, "Equalized interconnects for on-chip networks: Modeling and optimization framework," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, Nov. 2007, pp. 552–559.

[20] B. Kim and V. Stojanovic, "Characterization of equalized and repeated interconnects for NoC applications," *IEEE Design Test Comput.*, vol. 25, no. 5, pp. 430–439, Sep./Oct. 2008.

[21] H. Cho, P. Kapur, and K. C. Saraswat, "Power comparison between high-speed electrical and optical interconnects for interchip communication," *J. Lightw. Technol.*, vol. 22, no. 9, pp. 2021–2033, 2004.

[22] K.-H. Koo, H. Cho, P. Kapur, and K. C. Saraswat, "Performance comparisons between carbon nanotubes, optical, and cu for future high-performance on-chip interconnect applications," *IEEE Trans. Electron Devices*, vol. 54, pp. 3206–3215, 2007.

[23] H. Cho, K.-H. Koo, P. Kapur, and D. C. Saraswat, "Performance comparisons between Cu/low-κ, carbon-nanotube, and optics for future on-chip interconnects," *IEEE Electron Device Lett.*, vol. 29, pp. 122–124, 2008.

[24] J. H. Collet, F. Caignet, F. Sellaye, and D. Litaize, "Performance constraints for onchip optical interconnects," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 425–432, 2003.

[25] O. Kibar, D. A. Van Blerkom, C. Fan, and S. C. Esener, "Power minimization and technology for digital free-space optoelectronic interconnections," *J. Lightw. Technol.*, vol. 17, pp. 546–555, 1999.

[26] D. Huang, T. Sze, A. Landing, R. Lytel, and H. L. Davidson, "Optical interconnects: Out of the box forever?" *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 614–623, 2003.

[27] E. Berglind, L. Thylén, B. Jaskorzynska, and C. Svensson, "A comparison of dissipated power and signal-to-noise ratios in electrical and optical interconnects," *J. Lightw. Technol.*, vol. 17, pp. 68–73, 1999.

[28] C. Svensson, "Electrical interconnects revitalized," *IEEE Trans. Very Large Scale (VLSI) Syst.*, vol. 10, pp. 777–788, 2002.

[29] D. A. B. Miller, A. Bhatnagar, S. Palermo, A. Emami-Neyestanak, and M. A. Horowitz, "Opportunities for optics in integrated circuits applications," in *Proc. IEEE (ISSCC)*

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

Dig. Tech. Papers, vol. 1, San Francisco, CA, Feb. 6–10, 2005, pp. 86–87.

[30] C. Debaes, A. Bhatnagar, D. Agarwal, R. Chen, G. A. Keeler, N. C. Helman, H. Thienpont, and D. A. B. Miller, "Receiver-less optical clock injection for clock distribution networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 400–409, 2003.

[31] R. Urata, L. Y. Nathawad, R. Takahashi, K. Ma, D. A. B. Miller, B. A. Wooley, and J. S. Harris, "Photonic A/D conversion using low-temperature-grown GaAs MSM switches integrated with Si-CMOS," *J. Lightw. Technol.*, vol. 21, pp. 3104–3115, 2003.

[32] G. A. Keeler, B. E. Nelson, D. Agarwal, C. Debaes, N. C. Helman, A. Bhatnagar, and D. A. B. Miller, "The benefits of ultrashort optical pulses in optically-interconnected systems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 477–485, 2003.

[33] D. Agarwal, G. A. Keeler, C. Debaes, B. E. Nelson, N. C. Helman, and D. A. B. Miller, "Latency reduction in optical interconnects using short optical pulses," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 410–418, 2003.

[34] G. A. Keeler, B. E. Nelson, D. Agarwal, and D. A. B. Miller, "Skew and jitter removal using short optical pulses for optical interconnection," *IEEE Photon. Technol. Lett.*, vol. 12, pp. 714–716, 2000.

[35] B. E. Nelson, G. A. Keeler, D. Agarwal, N. C. Helman, and D. A. B. Miller, "Wavelength division multiplexed optical interconnect using short pulses," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 486–491, 2003.

[36] M. Horowitz, C.-K. Yang, and S. Sidiropoulos, *IEEE Micro*, vol. 18, pp. 250–27, Jan./Feb. 1998.

[37] P. M. Fauchet and J. H. Shen, Eds., "Issue on silicon photonics," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, no. 6, pt. 2, 2006.

[38] G. T. Reed and A. P. Knights, *Silicon Photonics—An Introduction*. Chichester, U.K.: Wiley, 2004.

[39] L. Pavesi and D. J. Lockwood, Eds., *Silicon Photonics*. Berlin, Germany: Springer-Verlag, 2004.

[40] N. Magen, A. Kolodny, U. Weiser, and N. Shamir, "Interconnect-power dissipation in a microprocessor," in *Proc. 2004 Int. Workshop System Level Interconnect Prediction (ACM, 2004) Session Interconnect Anal. SoCs Microprocess.*, pp. 7–13.

[41] L. A. Barroso, "The price of performance," *ACM Queue*, vol. 3, no. 7, pp. 48–53, Sep. 2005.

[42] J. G. Koomey, "Estimating total power consumption by servers in the U.S. and the world." [Online]. Available: http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf

[43] U.S. Environmental Protection Agency ENERGY STAR Program, "Report to Congress on server and data center energy efficiency, public law 109–431," Aug. 2, 2007. [Online]. Available: http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf

[44] L. A. Barroso and U. Hölzle, "The case of energy-proportional computing," *IEEE Computer*, vol. 40, pp. 33–37, Dec. 2007.

[45] Energy Information Administration, "Statistics on renewable and alternative fuels," 2007. [Online]. Available: http://www.eia.doe.gov/fuelrenewable.html

[46] "SMART 2020: Enabling the low carbon economy in the information age." [Online]. Available: http://www.gesi.org/index.php?article_id=43

[47] S. Palermo, A. Emami-Neyestanak, and M. Horowitz, "A 90 nm CMOS 16 Gb/s transceiver for optical interconnects," *IEEE J. Solid-State Circuits*, vol. 43, pp. 1235–1246, 2008.

[48] H. Hatamkhani, F. Lambrecht, V. Stojanovic, and C.-K. K. Yang, "Power-centric design of high-speed I/Os," in *Proc. 43rd ACM/IEEE Design Automa. Conf. 2006*, 2006, pp. 867–872.

[49] K.-L. J. Wong, A. Rylyakov, and C.-K. K. Yang, "A 5-mW 6-Gb/s quarter-rate sampling receiver with a 2-tap DFE using soft decisions," *IEEE J. Solid-State Circuits*, vol. 42, pp. 881–888, 2007.

[50] A. Emami-Neyestanak, A. Varzaghani, J. F. Bulzacchelli, A. Rylyakov, C.-K. K. Yang, and D. J. Friedman, "A 6.0-mW 10.0-Gb/s receiver with switched-capacitor summation DFE," *IEEE J. Solid-State Circuits*, vol. 42, pp. 889–896, 2007.

[51] R. J. Drost, R. D. Hopkins, R. Ho, and I. E. Sutherland, "Proximity communication," *IEEE J. Solid-State Circuits*, vol. 39, pp. 1529–1535, 2004.

[52] W. R. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A. M. Sule, M. Steer, and P. D. Franzon, "Demystifying 3D ICs: The pros and cons of going vertical," *IEEE Design Test Comput.*, vol. 22, pp. 498–510, Nov.-Dec. 2005.

[53] A. Fazzi, R. Canegallo, L. Ciccarelli, L. Magagni, F. Natali, E. Jung, P. Rolandi, and R. Guerrieri, "3-D capacitive interconnections with mono- and bi-directional capabilities," *IEEE J. Solid-State Circuits*, vol. 43, pp. 275–284, 2008.

[54] S. R. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, A. Singh, T. Jacob, S. Jain, V. Erraguntia, C. Roberts, Y. Hoskote, N. Borkar, and S. Borkar, "An 80-tile sub-100 W teraFLOPS processor in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 43, pp. 29–41, 2008.

[55] A. V. Krishnamoorthy, R. Ho, B. O'Krafka, J. E. Cunningham, J. Lexau, and X. Zheng, "Potentials of Group IV photonics interconnects for 'red-shift' computing applications," presented at the IEEE 2007 4th Int. Conf. Group IV Photonics, Tokyo, Japan, Sep. 19–21, 2007, Paper PLE2.1.

[56] T. H. Dunigan, Jr., J. S. Vetter, J. B. White, III, and P. H. Worley, "Performance evaluation of the Cray X1 distributed shared-memory architecture," *IEEE Micro*, vol. 25, pp. 30–40, Jan.-Feb. 2005.

[57] R. Drost, C. Forrest, B. Guenin, R. Ho, A. V. Krishnamoorthy, D. Cohen, J. E. Cunningham, B. Tourancheau, A. Zingher, A. Chow, G. Lauterbach, and I. Sutherland, "Challenges in building a flat-bandwidth memory hierarchy for a large-scale computer with proximity communication," in *Proc. IEEE 13th Symp. High Perform. Interconnects*, Aug. 17–19, 2005, pp. 13–22.

[58] G. Bell, J. Gray, and A. Szalay, "Petascale computational systems," *Computer*, vol. 39, no. 1, pp. 110–112, Jan. 2006.

[59] J. Gray and P. Shenoy, "Rules of thumb in data engineering," in *Proc. Int. Conf. Data Engineering (ICDE'00)*, 2000. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/ICDE.2000.839382

[60] J. D. Owens, W. J. Dally, R. Ho, D. N. Jayasimha, S. W. Keckler, and L.-S. Peh, "Research challenges for on-chip interconnection networks," *IEEE Micro*, pp. 96–108, Sep.–Oct. 2007.

[61] A. H. Gnauck, R. W. Tkach, A. R. Chraplyvy, and T. Li, "High-capacity optical transmission systems," *J. Lightw. Technol.*, vol. 26, pp. 1032–1045, 2008.

[62] A. H. Gnauck, G. Charlet, P. Tran, P. J. Winzer, C. R. Doerr, J. C. Centanni, E. C. Burrows, T. Kawanishi, T. Sakamoto, and K. Higuma, "25.6-Tb/s WDM transmission of polarization-multiplexed RZ-DQPSK signals," *J. Lightw. Technol.*, vol. 26, pp. 79–84, 2008.

[63] W. Bogaerts, P. Dumon, D. Van Thourhout, D. Taillaert, P. Jaenen, J. Wouters, S. Beckx, V. Wiaux, and R. G. Baets, "Compact wavelength-selective functions in silicon-on-insulator photonics wires," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1394–1401, 2006.

[64] S. Janz, P. Cheben, D. Dalacu, A. Dalage, A. Densmore, B. Lamontagne, R.-J. Picard, E. Post, J. H. Schmid, P. Waldron, C.-X. Xu, K. P. Yap, and W. N. Ye, "Microphotonic elements for integration on the silicon-on-insulator waveguide platform," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1402–1415, 2006.

[65] M. K. Smit, "New focusing and dispersive planar component based on an optical phased-array," *Electron. Lett.*, vol. 24, pp. 385–386, 1988.

[66] C. Dragone, "An N × N optical multiplexer using a planar arrangement of two star couplers," *IEEE Photon. Technol. Lett.*, vol. 3, pp. 812–815, 1991.

[67] H. Takahashi, K. Oda, H. Toba, and Y. Inoue, "Transmission characteristics of arrayed waveguide N × N wavelength multiplexer," *J. Lightw. Technol.*, vol. 13, pp. 447–455, 1995.

[68] K. Jia, J. Yang, Y. Hao, X. Jiang, M. Wang, W. Wang, Y. Wu, and Y. Wang, "Turning-mirror-integrated arrayed-waveguide gratings on silicon-on-insulator," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1329–1334, 2006.

[69] S. Zheng, H. Chen, and A. W. Poon, "Microring-resonator cross-connect filters in silicon nitride: Rib waveguide dimensions dependence," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1380–1387, 2006.

[70] B. G. Lee, B. A. Small, Q. Xu, M. Lipson, and K. Bergman, "Characterization of a 4 × 4 Gb/s parallel electronic bus to WDM optical link silicon photonic translator," *IEEE Photon. Technol. Lett.*, vol. 19, pp. 456–459, 2007.

[71] F. Horst, W. M. J. Green, B. J. Offrein, and Y. Vlasov, "Echelle grating WDM (de-)multiplexers in SOI technology based on a design with two stigmatic points," in *Proc. SPIE*, 2008, vol. 6996, p. 69960R.

[72] J. Brouckaert, W. Bogaerts, S. Sevaraja, P. Dumon, R. Baets, and D. Van Thourhout, "Planar concave grating demultiplexer with high reflective bragg reflector facets," *IEEE Photon. Technol. Lett.*, vol. 20, pp. 309–311, 2008.

[73] R. Zengerle, "Light propagation in singly and doubly periodic planar waveguides," *J. Modern Opt.*, vol. 34, no. 12, pp. 1589–1617, 1987.

[74] H. Kosaka, T. Kawashima, A. Tomita, M. Notomi, T. Tamamura, T. Sata, and S. Kawakami, "Photonic crystals for micro lightwave circuits using wavelength-

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

dependent angular beam steering," *Appl. Phys. Lett.*, vol. 74, pp. 1370–1372, 1999.

[75] B. Momeni and A. Adibi, "Preconditioned superprism-based photonic crystal demultiplexers: Analysis and design," *Appl. Opt.*, vol. 45, pp. 8466–8476, 2006.

[76] L. J. Wu, M. Mazilu, T. Karle, and T. F. Krauss, "Superprism phenomena in planar photonic crystals," *IEEE J. Quantum Electron.*, vol. 38, pp. 915–918, 2002.

[77] T. Baba and M. Nakamura, "Photonic crystal light deflection devices using the superprism effect," *IEEE J. Quantum Electron.*, vol. 38, pp. 909–914, 2002.

[78] B. E. Nelson, M. Gerken, D. A. B. Miller, R. Piestun, C.-C. Lin, and J. S. Harris, Jr., "Use of a dielectric stack as a one-dimensional photonic crystal for wavelength demultiplexing by beam shifting," *Opt. Lett.*, vol. 25, pp. 1502–1504, 2000.

[79] M. Gerken and D. A. B. Miller, "Multilayer thin-film structures with high spatial dispersion," *Appl. Opt.*, vol. 42, pp. 1330–1345, 2003.

[80] M. Gerken and D. A. B. Miller, "Wavelength demultiplexer using the spatial dispersion of multilayer thin-film structures," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 1097–1099, 2003.

[81] M. Gerken and D. A. B. Miller, "Multilayer thin-film stacks with steplike spatial beam shifting," *J. Lightw. Technol.*, vol. 22, pp. 612–618, 2004.

[82] M. Gerken and D. A. B. Miller, "Limits to the performance of dispersive thin-film stacks," *Appl. Opt.*, vol. 44, no. 18, pp. 3349–3357, 2005.

[83] M. Gerken and D. A. B. Miller, "The relationship between the superprism effect in one-dimensional photonic crystals and spatial dispersion in non-periodic thin-film stacks," *Opt. Lett.*, vol. 30, no. 18, pp. 2475–2477, Sep. 15, 2005.

[84] D. A. B. Miller, "Fundamental limit for optical components," *J. Opt. Soc. Amer. B*, vol. 24, pp. A1–A18, 2007.

[85] D. A. B. Miller, "Fundamental limit to linear one-dimensional slow light structures," *Phys. Rev. Lett.*, vol. 99, p. 203903, 2007.

[86] C. Gunn, "CMOS photonics for high-speed interconnects," *IEEE Micro*, vol. 26, pp. 58–66, Mar.–Apr. 2006.

[87] D. Van Thourhout, G. Roelkens, R. Baets, W. Bogaerts, J. Brouckaert, P. Debackere, P. Dumon, S. Scheerlinck, J. Schrauwen, D. Taillaert, F. Van Laere, and J. Van Campenhout, "Coupling mechanisms for a heterogeneous silicon nanowire platform," *Semicond. Sci. Technol.*, vol. 23, p. 064004, 2008, (9pp).

[88] V. R. Almeida, R. R. Panepucci, and M. Lipson, "Nanotaper for compact mode conversion," *Opt. Lett.*, vol. 28, pp. 1302–1304, 2003.

[89] X. Zheng, J. E. Cunningham, I. Shubin, J. Simons, M. Asghari, D. Feng, H. Lei, D. Zheng, H. Liang, C. Kung, J. Luff, T. Sze, D. Cohen, and A. V. Krishnamoorthy, "Optical proximity communication using reflective mirrors," *Opt. Express*, vol. 16, pp. 15052–15058, 2008.

[90] L. Vivien, X. Le Roux, S. Laval, E. Cassan, and D. Marris-Morini, "Design, realization, and characterization of 3-D taper for fiber/micro-waveguide coupling," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1354–1358, 2006.

[91] G. Masanovic, G. Reed, W. Headley, B. Timotijevic, V. Passaro, R. Atta, G. Ensell, and A. Evans, "A high efficiency input/output coupler for small silicon photonic devices," *Opt. Express*, vol. 13, pp. 7374–7379, 2005.

[92] G. Z. Mashanovich, M. Milosevic, P. Matavulj, S. Stankovic, B. Timotijevic, P. Y. Yang, E. J. Teo, M. B. H. Breese, A. A. Bettiol, and G. T. Reed, "Silicon photonic waveguides for different wavelength regions," *Semicond. Sci. Technol.*, vol. 23, p. 064002, 2008, (9pp).

[93] W. Bogaerts, R. Baets, P. Dumon, V. Wiaux, S. Beckx, D. Taillaert, B. Luyssaert, J. Van Campenhout, P. Bienstman, and D. Van Thourhout, "Nanophotonic waveguides in silicon-on-insulator fabricated with CMOS technology," *J. Lightw. Technol.*, vol. 23, pp. 401–412, 2005.

[94] D. K. Sparacin, S. J. Spector, and L. C. Kimerling, "Silicon waveguide sidewall smoothing by wet chemical oxidation," *J. Lightw. Technol.*, vol. 23, pp. 2455–2461, 2005.

[95] H. Yamada, T. Chu, S. Ishida, and Y. Arakawa, "Si photonic wire waveguide devices," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1371–1379, 2006.

[96] E. Cassan, S. Laval, S. Lardenois, and A. Koster, "On-chip optical interconnects with compact and low-loss light distribution in silicon-on-insulator rib waveguides," *IEEE J. Sel. Top. Quantum Electron.*, vol. 9, pp. 460–464, 2003.

[97] N. Streibl, K.-H. Brenner, A. Huang, J. Jahns, J. Jewell, A. W. Lohmann, D. A. B. Miller, M. Murdocca, M. E. Prise, and T. Sizer, "Digital optics," *Proc. IEEE*, vol. 77, pp. 1954–1969, 1989.

[98] F. B. McCormick, T. J. Cloonan, F. A. P. Tooley, A. L. Lentine, J. M. Sasian, J. L. Brubaker, R. L. Morrison, S. L. Walker, R. J. Crisci, R. A. Novotny, S. J. Hinterlong, H. S. Hinton, and E. Kerbis, "Six-stage digital free-space optical switching network using symmetric self-electro-optic-effect devices," *Appl. Opt.*, vol. 32, pp. 5153–5171, 1993.

[99] F. B. McCormick, T. J. Cloonan, A. L. Lentine, J. M. Sasian, R. L. Morrison, M. G. Beckman, S. L. Walker, M. J. Wojcik, S. J. Hinterlong, R. J. Crisci, R. A. Novotny, and H. S. Hinton, "Five-stage free-space optical switching network with field-effect transistor self-electro-optic-effect-device smart-pixel arrays," *Appl. Opt.*, vol. 33, pp. 1601–1618, 1994.

[100] R. Barbieri, P. Benabes, T. Bierhoff, J. J. Caswell, A. Gauthier, J. Jahns, M. Jarczynski, P. Lukowicz, J. Oksman, G. A. Russell, J. Schrage, J. F. Snowdon, O. Stübbe, G. Troster, and M. Wirz, "Design and construction of the high-speed optoelectronic memory system demonstrator," *Appl. Opt.*, vol. 47, pp. 3500–3512, 2008.

[101] M. B. Venditti, E. Laprise, J. Faucher, P.-O. Laprise, J. Eduardo, A. Lugo, and D. V. Plant, "Design and test of an optoelectronic-VLSI chip with 540-element receiver-transmitter arrays using differential optical signaling," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 361–379, 2003.

[102] A. C. Walker, M. P. Y. Desmulliez, M. G. Forbes, S. J. Fancey, G. S. Buller, M. R. Taghizadeh, J. A. B. Dines, C. R. Stanley, G. Pennelli, A. R. Boyd, P. Horan, D. Byrne, J. Hegarty, S. Eitel, H. P. Gauggel, K. H. Gulden, A. Gauthier, P. Benabes, J. L. Gutzwiller, and M. Goetz, Desmulliez MPY, "Design and construction of an optoelectronic crossbar switch containing a terabit per second free-space optical interconnect," *IEEE J. Sel. Topics Quantum Electron.*, vol. 5, pp. 236–249, 1999.

[103] D. V. Plant and A. G. Kirk, "Optical interconnects at the chip and board level: Challenges and solutions," *Proc. IEEE*, vol. 88, pp. 806–818, 2000.

[104] M. W. Haney, M. P. Christensen, P. Milojkovic, G. J. Fokken, M. Vickberg, B. K. Gilbert, J. Rieve, J. Ekman, P. Chandramani, and F. Kiamilev, "Description and evaluation of the fast-net smart pixel-based optical interconnection prototype," *Proc. IEEE*, vol. 88, pp. 819–828, 2000.

[105] K. W. Goossen, J. A. Walker, L. A. D'Asaro, B. Tseng, R. Leibenguth, D. Kossives, D. D. Bacon, D. Dahringer, L. M. F. Chirovsky, A. L. Lentine, and D. A. B. Miller, "GaAs MQW modulators integrated with silicon CMOS," *IEEE Photon. Technol. Lett.*, vol. 7, pp. 360–362, 1995.

[106] A. V. Krishnamoorthy and K. W. Goossen, "Optoelectronic-VLSI: Photonics integrated with VLSI circuits," *IEEE J. Sel. Topics Quantum Electron.*, vol. 4, pp. 899–912, 1998.

[107] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, "Electric field dependence of optical absorption near the bandgap of quantum well structures," *Phys. Rev.*, vol. B32, pp. 1043–1060, 1985.

[108] M. P. Christensen, P. Milojkovic, M. J. McFadden, and M. W. Haney, "Multiscale optical design for global chip-to-chip optical interconnections and misalignment tolerant packaging," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 548–556, 2003.

[109] M. Jarczynski, T. Seiler, and J. Jahns, "Integrated three-dimensional optical multilayer using free-space optics," *Appl. Opt.*, vol. 45, pp. 6335–6341, 2006.

[110] C. Debaes, M. Vervaeke, V. Baukens, H. Ottevaere, P. Vynck, P. Tuteleers, B. Volckaerts, W. Meeus, M. Brunfaut, J. Van Campenhout, A. Hermanne, and H. Thienpont, "Low-cost microoptical modules for MCM level optical interconnections," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, pp. 518–530, 2003.

[111] R. Zia, M. D. Selker, P. B. Catrysse, and M. L. Brongersma, "Geometries and materials for subwavelength surface plasmon modes," *J. Opt. Soc. Amer. A*, vol. 21, no. 12, pp. 2442–2446, 2004.

[112] R. Zia, J. A. Schuller, and M. A. Brongersma, "Near-field characterization of guided polariton propagation and cutoff in surface plasmon waveguides," *Phys. Rev. B*, vol. 74, p. 165415, 2006.

[113] N.-N. Feng, M. L. Brongersma, and L. Dal Negro, "Metal-dielectric slot-waveguide structures for the propagation of surface plasmon polaritons at 1.55 μm," *IEEE J. Quantum Electron.*, vol. 43, no. 6, pp. 479–485, 2007.

[114] J. A. Dionne, L. A. Sweatlock, H. A. Atwater, and A. Polman, "Plasmon slot waveguides: Towards chip-scale propagation with subwavelength-scale localization," *Phys. Rev. B*, vol. 73, p. 035407, 2006.

[115] G. Veronis and S. H. Fan, "Modes of subwavelength plasmonic slot waveguides," *J. Lightw. Technol.*, vol. 25, pp. 2511–2521, 2007.

[116] D.-S. Ly-Gagnon, S. E. Kocabas, and D. A. B. Miller, "Characteristic impedance model for plasmonic metal slot waveguides,"

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

IEEE J. Sel. Topics Quantum Electron., vol. 14, pp. 1473–1478, 2008.

[117] S. E. Kocabas, G. Veronis, D. A. B. Miller, and S. H. Fan, "Transmission line and equivalent circuit models for plasmonic waveguide components," *IEEE J. Sel. Topics Quantum Electron.*, vol. 14, pp. 1462–1472, 2008.

[118] R. S. Tucker, "Energy consumption in digital optical ICs with plasmon waveguide interconnects," *IEEE Photon. Technol. Lett.*, vol. 19, pp. 2036–2038, 2007.

[119] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm early design exploration," *IEEE Trans. Electron Devices*, vol. 53, pp. 2816–2823, 2006.

[120] Y. Cao, T. Sato, M. Orshansky, D. Sylvester, and C. Hu, "New paradigm of predictive MOSFET and interconnect modeling for early circuit design," in *Proc. IEEE CICC*, Jun. 2000, pp. 201–204.

[121] A. V. Krishnamoorthy and D. A. B. Miller, "Scaling Optoelectronic-VLSI circuits into the 21st century: A technology roadmap," *IEEE J. Sel. Topics Quantum Electron.*, vol. 2, no. 1, pp. 55–76, 1996.

[122] A. K. Okyay, A. J. Pethe, D. Kuzum, S. Latif, D. A. B. Miller, and K. C. Saraswat, "SiGe optoelectronic metal-oxide semiconductor field-effect transistor," *Opt. Lett.*, vol. 32, pp. 2022–2024, 2007.

[123] A. K. Okyay, D. Kuzum, S. Latif, D. A. B. Miller, and K. C. Saraswat, "Silicon germanium CMOS optoelectronic switching device: Bringing light to latch," *IEEE Trans. Electron Devices*, vol. 54, pp. 3252–3259, 2007.

[124] S. Sahni, X. Luo, J. Liu, Y. Xie, and E. Yablonovitch, "Junction field-effect-transistor-based germanium photodetector on silicon-on-insulator," *Opt. Lett.*, vol. 33, pp. 1138–1140, 2008.

[125] D. Ahn, C. Hong, J. Liu, W. Giziewicz, M. Beals, L. C. Kimerling, J. Michel, J. Chen, and F. X. Kärtner, "High performance, waveguide integrated Ge photodetectors," *Opt. Express*, vol. 15, pp. 3916–3921, 2007.

[126] Z. Huang, N. Kong, X. Guo, M. Liu, N. Duan, A. L. Beck, S. K. Banerjee, and J. C. Campbell, "21-GHz-bandwidth germanium-on-silicon photodiode using thin SiGe buffer layers," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1450–1454, 2006.

[127] A. Nemecek, G. Zach, S. Swoboda, K. Oberhauser, and H. Zimmermann, "Integrated BiCMOS p-i-n photodetectors with high bandwidth and high responsivity," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1469–1475, 2006.

[128] S. J. Koester, J. D. Schaub, G. Dehlinger, and J. O. Chu, "Germanium-on-SOI infrared detectors for integrated photonic applications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1489–1502, 2006.

[129] L. Colace, M. Balbi, G. Masini, G. Assanto, H. C. Luan, and L. C. Kimerling, "Ge on Si p-i-n photodiodes operating at 10 Gb/s," *Appl. Phys. Lett.*, vol. 88, p. 101111, 2006.

[130] O. Fidaner, A. K. Okyay, J. E. Roth, R. K. Schaevitz, Y.-H. Kuo, K. C. Saraswat, J. S. Harris, Jr., and D. A. B. Miller, "Ge-SiGe quantum-well waveguide photodetectors on silicon for the near-infrared," *IEEE Photon. Technol. Lett.*, vol. 19, pp. 1631–1633, 2007.

[131] Y. Kang, M. Zadka, S. Litski, G. Sarid, M. Morse, M. J. Paniccia, Y. Kuo, J. Bowers, A. Beling, H. Liu, D. C. McIntosh, J. Campbell, and A. Pauchard, "Epitaxially-grown Ge/Si avalanche photodiodes for 1.3 μm light detection," *Opt. Express*, vol. 16, pp. 9365–9371, 2008.

[132] L. Chen, P. Dong, and M. Lipson, "High performance germanium photodetectors integrated on submicron silicon waveguides by low temperature wafer bonding," *Opt. Express*, vol. 16, pp. 11513–11518, 2008.

[133] T. Ishi, J. Fujikata, K. Makita, T. Baba, and K. Ohashi, "Si nano-photodiode with a surface plasmon antenna," *Jpn. J. Appl. Phys.*, vol. 44, no. 12, pp. L364–L366, 2005.

[134] L. Tang, D. A. B. Miller, A. K. Okyay, J. A. Matteo, Y. Yuen, K. C. Saraswat, and L. Hesselink, "C-shaped nanoaperture-enhanced germanium photodetector," *Opt. Lett.*, vol. 31, pp. 1519–1521, 2006.

[135] L. Tang, S. E. Kocabas, S. Latif, A. K. Okyay, D.-S. Ly-Gagnon, K. C. Saraswat, and D. A. B. Miller, "Nanometre-scale germanium photodetector enhanced by a near-infrared dipole antenna," *Nature Photon.*, vol. 2, pp. 226–229, 2008.

[136] L. Tang, S. Latif, and D. A. B. Miller, "Plasmonic device in CMOS," presented at the 21st Annu. Lasers Electro Optics Soc. Meeting (LEOS 2008), Newport Beach, CA, Nov. 2008, Paper TuA2.

[137] Y.-C. Chang and L. A. Coldren, "Optimization of VCSEL structure for high-speed operation," in *Proc. IEEE 21st ISLC*, Sorrento, Italy, Sep. 14–18, 2008, pp. 159–160.

[138] D. Englund, H. Altug, B. Ellis, and J. Vuckovic, "Ultrafast photonic crystal lasers," *Laser Photon. Rev.*, vol. 2, pp. 264–274, 2008.

[139] J. van Eisden, M. Yakimov, V. Tokranov, M. Varanasi, E. M. Mohammed, I. A. Young, and S. R. Oktyabrsky, "Optically decoupled loss modulation in a duo-cavity VCSEL," *IEEE Photon. Technol. Lett.*, vol. 20, pp. 42–44, 2008.

[140] L. C. Chuang, M. Moewe, C. Chase, N. P. Kobayashi, C. Chang-Hasnain, and S. Crankshaw, "Critical diameter for III-V nanowires grown on lattice-mismatched substrates," *Appl. Phys. Lett.*, vol. 90, p. 043115, 2007.

[141] A. W. Fang, H. Park, O. Cohen, R. Jones, M. J. Paniccia, and J. E. Bowers, "Electrically pumped hybrid AlGaInAs-silicon evanescent laser," *Opt. Express*, vol. 14, pp. 9203–9210, 2006.

[142] A. W. Fang, B. R. Koch, K. Gan, H. Park, R. Jones, O. Cohen, M. J. Paniccia, D. J. Blumenthal, and J. E. Bowers, "A racetrack mode-locked silicon evanescent laser," *Opt. Express*, vol. 16, pp. 1393–1398, 2008.

[143] S. G. Cloutier, P. A. Kossyrev, and J. Xu, "Optical gain & stimulated emission in periodic nanopatterned crystalline silicon," *Nat. Mater.*, vol. 4, p. 887, 2005.

[144] J. Liu, X. Sun, D. Pan, X. Wang, L. C. Kimerling, T. L. Koch, and J. Michel, "Tensile-strained, n-type Ge as a gain medium for monolithic laser integration on Si," *Opt. Express*, vol. 15, pp. 11272–11277, 2007.

[145] P. J. Delfyett, S. Gee, M.-T. Choi, H. Izadpanah, W. Lee, S. Ozharar, F. Quinlan, and T. Yilmaz, "Optical frequency combs from semiconductor lasers and applications in ultrawideband signal processing and communications," *J. Lightw. Technol.*, vol. 24, pp. 2701–2719, 2006.

[146] J. S. Weiner, D. A. B. Miller, and D. S. Chemla, "Quadratic electro-optic effect due to the quantum-confined stark effect in quantum wells," *Appl. Phys. Lett.*, vol. 50, pp. 842–844, 1987.

[147] A. Liu, L. Liao, D. Rubin, J. Basak, Y. Chetrit, H. Nguyen, R. Cohen, N. Izhaky, and M. Paniccia, "Recent development in a high-speed silicon optical modulator based on reverse-biased pn diode in a silicon waveguide," *Semicond. Sci. Technol.*, vol. 23, p. 064001, 2008.

[148] R. A. Soref and B. R. Bennett, "Electrooptical effects in silicon," *IEEE J. Quantum Electron.*, vol. QE-23, no. 1, pp. 123–129, 1987.

[149] W. M. Green, M. J. Rooks, L. Sekaric, and Y. A. Vlasov, "Ultra-compact, low RF power, 10 Gb/s silicon Mach-Zehnder modulator," *Opt. Express*, vol. 15, pp. 17106–17113, 2007.

[150] L. Gu, W. Jiang, X. Chen, and R. T. Chen, "Physical mechanism of p-i-n-diode-based photonic crystal silicon electrooptic modulators for gigahertz operation," *IEEE J. Sel. Topics Quantum Electron.*, vol. 14, pp. 1132–1139, 2008.

[151] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson, "12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators," *Opt. Express*, vol. 15, pp. 430–436, 2007.

[152] Q. Xu, D. Fattal, and R. G. Beausoleil, "Silicon microring resonators with 1.5-μm radius," *Opt. Express*, vol. 16, pp. 4309–4315, 2008.

[153] M. Lipson, "Compact electro-optic modulators on a silicon chip," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1520–1526, 2006.

[154] Y. Li, L. Zhang, M. Song, B. Zhang, J. Yang, R. G. Beausoleil, A. E. Willner, and P. D. Dapkus, "Coupled-ring-resonator-based silicon modulator for enhanced performance," *Opt. Express*, vol. 16, pp. 13342–13348, 2008.

[155] D. A. B. Miller, D. S. Chemla, and S. Schmitt-Rink, "Relation between electroabsorption in bulk semiconductors and in quantum wells: The quantum-confined Franz-Keldysh effect," *Phys. Rev.*, vol. B33, pp. 6976–6982, 1986.

[156] Y. Luo, B. Xiong, J. Wang, P. F. Cai, and C. Z. Sun, "40 GHz AlGaInAs multiple-quantum-well integrated electroabsorption modulator/distributed feedback laser based on identical epitaxial layer scheme," *Jpn. J. Appl. Phys.*, vol. 45, *Lett. Express Lett.*, pt. 2, pp. L1071–L1073, 2006.

[157] S. Schmitt-Rink, D. S. Chemla, W. H. Knox, and D. A. B. Miller, "How fast is excitonic electroabsorption?" *Opt. Lett.*, vol. 15, pp. 60–62, 1990.

[158] M. B. Yairi, C. W. Coldren, D. A. B. Miller, and J. S. Harris, Jr., "High-speed, optically-controlled surface-normal modulator based on diffusive conduction," *Appl. Phys. Lett.*, vol. 75, no. 5, pp. 597–599, 1999.

[159] M. Chacinski, U. Westergren, B. Willen, J. Stoltz, and L. Thylen, "Electroabsorption modulators suitable for 100 Gb/s ethernet," *IEEE Electron Device Lett.*, vol. 29, pp. 1014–1016, 2008.

[160] S. Kodama, T. Yoshimatsu, and H. Ito, "500 Gbit/s optical gate monolithically integrating photodiode and electroabsorption modulator," *Electron. Lett.*, vol. 40, pp. 555–556, 2004.

[161] N. C. Helman, J. E. Roth, D. P. Bour, H. Altug, and D. A. B. Miller, "Misalignment-tolerant surface-normal low-voltage modulator for optical

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.

interconnects," *IEEE J. Sel. Topics Quantum Electron.*, vol. 11, pp. 338–342, Mar./Apr. 2005.

[162] J. E. Roth, O. Fidaner, R. K. Schaevitz, Y.-H. Kuo, T. I. Kamins, J. S. Harris, and D. A. B. Miller, "Optical modulator on silicon employing germanium quantum wells," *Opt. Express*, vol. 15, pp. 5851–5859, 2007.

[163] J. E. Roth, O. Fidaner, E. H. Edwards, R. K. Schaevitz, Y.-H. Kuo, N. C. Helman, T. I. Kamins, J. S. Harris, and D. A. B. Miller, "C-band side-entry Ge quantum-well electroabsorption modulator on SOI operating at 1 V swing," *Electron. Lett.*, vol. 44, pp. 49–50, 2008.

[164] K. W. Goossen, G. D. Boyd, J. E. Cunningham, W. Y. Jan, D. A. B. Miller, D. S. Chemla, and R. M. Lum, "GaAs-AlGaAs multiquantum well reflection modulators grown on GaAs and silicon substrates," *IEEE Photon. Technol. Lett.*, vol. 1, pp. 304–306, 1989.

[165] Y. Kuo, H. Chen, and J. E. Bowers, "High speed hybrid silicon evanescent electroabsorption modulator," *Opt. Express*, vol. 16, pp. 9936–9941, 2008.

[166] Y.-H. Kuo, Y.-K. Lee, Y. Ge, S. Ren, J. E. Roth, T. I. Kamins, D. A. B. Miller, and J. S. Harris, "Strong quantum-confined Stark effect in germanium quantum-well structures on silicon," *Nature*, vol. 437, pp. 1334–1336, 2005.

[167] Y.-H. Kuo, Y. K. Lee, Y. Ge, S. Ren, J. E. Roth, T. I. Kamins, D. A. B. Miller, and J. S. Harris, Jr., "Quantum-confined stark effect in Ge/SiGe quantum wells on Si for optical modulators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1503–1513, 2006.

[168] R. K. Schaevitz, J. E. Roth, S. Ren, O. Fidaner, and D. A. B. Miller, "Material properties in Si-Ge/Ge quantum wells," *IEEE J. Sel. Topics Quantum Electron.*, vol. 14, pp. 1082–1089, 2008.

[169] J. Liu, M. Beals, A. Pomerene, S. Bernardis, R. Sun, J. Cheng, L. C. Kimerling, and J. Michel, "Waveguide-integrated, ultralow-energy GeSi electro-absorption modulators," *Nature Photon.*, vol. 2, pp. 433–437, 2008.

ABOUT THE AUTHOR

David A. B. Miller (Fellow, IEEE) received the B.Sc. degree from St. Andrews University, Scotland, U.K., and the Ph.D. degree from Heriot-Watt University, Scotland, U.K., in 1979, both in physics.

He was with Bell Laboratories from 1981 to 1996. From 1987, he was a Department Head, latterly of the Advanced Photonics Research Department. He is currently the W. M. Keck Professor of Electrical Engineering at Stanford University, Stanford, CA, Director of the Solid State and Photonics Laboratory at Stanford, and a Co-Director of the Stanford Photonics Research Center. He also was Director of the Ginzton Laboratory at Stanford from 1997 to 2006. His research interests include nanophotonic and quantum-confined optoelectronic physics and devices, and fundamentals and applications of optics in information sensing, switching, and processing. He has published more than 220 scientific papers and delivered more than 130 conference invited talks. He has received 67 patents. He is the author of *Quantum Mechanics for Scientists and Engineers* (Cambridge, U.K.: Cambridge University Press, 2008). He has served on boards for various photonics companies.

Dr. Miller is a Fellow of the Royal Societies of London and Edinburgh, Optical Society of America (OSA), and APS and a member of the National Academy of Sciences. He has served as a Board Member of the OSA and the IEEE Lasers and Electro-Optics Society (LEOS), and in various other society and conference committees. He was President of LEOS in 1995. He received the Adolph Lomb Medal and the R. W. Wood Prize from the OSA, the International Prize in Optics from the International Commission for Optics, and the IEEE Third Millennium Medal. He has received honorary degrees from the Vrije Universiteit Brussel and Heriot-Watt University.

Authorized licensed use limited to: Stanford University. Downloaded on June 17, 2009 at 11:45 from IEEE Xplore. Restrictions apply.