Rationale and Challenges for Optical Interconnects to Electronic Chips

DAVID A. B. MILLER, FELLOW, IEEE

Invited Paper

The various arguments for introducing optical interconnections to silicon CMOS chips are summarized, and the challenges for optical, optoelectronic, and integration technologies are discussed. Optics could solve many physical problems of interconnects, including precise clock distribution, system synchronization (allowing larger synchronous zones, both on-chip and between chips), bandwidth and density of long interconnections, and reduction of power dissipation. Optics may relieve a broad range of design problems, such as crosstalk, voltage isolation, wave reflection, impedance matching, and pin inductance. It may allow continued scaling of existing architectures and enable novel highly interconnected or high-bandwidth architectures. No physical breakthrough is required to implement dense optical interconnects to silicon chips, though substantial technological work remains. Cost is a significant barrier to practical introduction, though revolutionary approaches exist that might achieve economies of scale. An Appendix analyzes scaling of on-chip global electrical interconnects, including line inductance and the skin effect, both of which impose significant additional constraints on future interconnects.

Keywords**—***Off-chip wiring, on-chip wiring, optical interconnects, quantum-well modulator, vertical-cavity surface-emitting laser.*

I. INTRODUCTION

Digital processing of information requires nonlinear devices and circuits for logical functions and storage, and also interconnections to carry the information from one place to another. The continuing exponential reduction in feature sizes on electronic chips, known as Moore's law [1], leads to ever larger numbers of faster devices at lower cost per device. This evolution is shifting the balance between devices and interconnection in digital processing systems; electrical interconnections, at least as we know them today, do not scale to keep up with the devices.

Problems with scaling electrical interconnections have been known for some time, at least implicitly. For example, essentially all telecommunications has moved away from electrical lines for long-distance traffic because the loss at high frequencies in electrical wires is too high. Also, inside computer systems, the buses that carry information from one part of the system to another run at rates much slower than the clock rate on the chips because of a variety of problems with electrical interconnects, including wave reflections on the lines. The existence of interconnect scaling problems has been highlighted recently because of the roadmaps created by the Semiconductor Industry Association (SIA) [1]. These roadmaps show that even on semiconductor chips themselves, where interconnects are short, plentiful, and inexpensive by any absolute measure, the global interconnects will become very difficult. It is already the case for electrical interconnections between chips that the performance is dominated by the interconnection medium rather than the devices at either end; sometime in the next decade, this will also be the case for many connections on chips.

There are several possible approaches to such interconnection scaling problems, and likely all of these will be used to some degree. Architectures could be changed to minimize interconnection. Design approaches could put increasing emphasis on the interconnection layout. Signaling on wires could be significantly improved though the use of a variety of techniques, such as equalization [2]–[4]. Most important for this discussion is a fourth approach—changing the physical means of interconnection. Optics is arguably a very interesting and different physical approach to interconnection that can in principle address most, if not all, of the problems encountered in electrical interconnections.

It should be emphasized here that the difficulties of electrical interconnections are not simply ones of scaling of the raw capacity of the interconnection system. There are a variety of other difficulties that are not so readily quantified

Manuscript received October 22, 1999; revised February 29, 2000. This work was supported by the Interconnect Focus Research Center (funded by the Semiconductor Industry Association and DARPA through MARCO), by the Semiconductor Research Corp., by DARPA under Grant MDA972-98-1-0002, and by the University of New Mexico.

The author is with Ginzton Laboratory, Stanford University, Stanford, CA 94305-4085 USA.

Publisher Item Identifier S 0018-9219(00)05247-6.

0018–9219/00$10.00 © 2000 IEEE

but that in the end could be dominant reasons for changing to a radical solution like optics. This includes issues such as voltage isolation, timing accuracy, and overall ease of design. The design issue might turn out to be an important one; for example, an electrical bus designed for 500 MHz may well not work at 600 MHz because of different loss, inductance, crosstalk, and wave reflection phenomena. By contrast, an optical system designed for 500 MHz might continue to work equally well up to 500 GHz (if we had the devices to drive and receive the optical signals) because the frequency of modulation has essentially no effect on the propagation of the light signals.

There are other possible physical approaches to improving electrical interconnections, including cooling the chips and/or circuits (to get lower resistance in lines), e.g., to 77 K, or using superconducting lines. Cooling to low room temperatures is already implemented in some computers. Cryogenic cooling would be physically possible with current technology. Superconducting materials are still not available for room temperature, and so the use of superconductors would also require significant cooling; unless temperatures <77 K are used, relatively novel, practical, high-temperature superconductor materials would have to be developed. The number of metal levels can be increased, with seven levels apparently currently in production, and larger numbers of levels under development. There is significant cost to developing further levels, and it is not clear that this is a scalable solution to interconnect problems in the long run. It is also possible to consider using off-chip wiring layers attached to the chip to augment the on-chip wiring. There is, however, an underlying scaling difficulty with electrical interconnections (as discussed below in Section II-B1), which limits both on-chip and off-chip wiring even if we have the ability to make many layers. One interesting approach is to stack chips in a three-dimensional (3-D) structure with appropriate vertical connections. This approach may well help, though power dissipation can become a problem in such approaches because the surface area is not increasing significantly as chips are stacked. This power dissipation increase may to some extent be offset by the opportunity to use shorter interconnections in such a 3-D structure. Even such 3-D structures do not avoid some of the underlying scaling limits of electrical interconnections. In addition, all of these electrical approaches do not address the other qualitative problems like voltage isolation, timing accuracy, and ease of design. Thus, in general, though there are other approaches to electrical interconnects that may well help, there are underlying scaling issues and other physical problems that remain.

Implementing optical interconnects to chips would also face many technical challenges. If we wish seriously to impact interconnections on-chip or chip-to-chip, we need to be considering technologies that can allow "dense" optical interconnects at the chip level, by which we mean at least hundreds or more likely thousands or more of optical interconnects for each chip. Without such numbers, most off-chip interconnects and long on-chip interconnects would have to remain electrical. Much sophisticated optical and optoelectronic technology has been developed for long-distance communications, but the requirements of dense interconnects are substantially different. Low power dissipation, small latency, small physical size, and the ability to integrate with mainstream silicon electronics in large numbers are all required for dense interconnects at the chip-to-chip or on-chip level. Existing optical telecommunications applications do not require any of these constraints, and the technologies developed do not satisfy them. Additionally, the discrete approaches used for long distances are likely not to be viable for dense interconnects. There are, however, other opportunities in optical and optoelectronic technology that have been researched over the last several years that are apparently capable of operating at the densities needed, though the technologies are often quite different from those of long-distance communications and are much less mature.

In this paper, we discuss first, in Section II, the potential benefits of optical interconnects, grouping these under scaling, timing, design simplification, architectural, and other benefits. Then in Section III we discuss some of the challenges for optics, and finish in Section IV with conclusions. An Appendix extends some scaling models for on-chip electrical interconnections to allow some conclusions about future on-chip electrical wiring.

II. Potential Benefits of Optics

A. Historical Background

Optical interconnects to electronic chips have been the subject of research for at least the last 15 years, starting with the seminal paper of Goodman *et al.* [5]. Since that time, a body of work (see, for example, [4] and [6]–[33]) has addressed potential benefits and limits of optics for interconnection [4], [7], [8], [12], [14]–[16], [20], [23], [24], [28], analysis of the relative benefits of optics versus electronics [4], [6], [9]–[11], [13], [21], [22], [26], [27], [30], [31], [33], and comparison of different kinds of optical approaches against one another [17]–[19], [25], [29]. Several of these papers review parts of this work (e.g., [20], [22], [27], and [28]).

In parallel with much of the work on optical interconnects, there was effort on optoelectronic digital computing. This evolved from work on optical switching devices toward optoelectronic switching device arrays, which in turn evolved into "smart pixels"—arrays of optoelectronic units with significant electronic processing between optical inputs and outputs [34], [35]. The emergence of electrically controlled optical modulators based on novel effects in quantum-well structures [36], including the family of devices known as self-electrooptic-effect devices (SEED's) [37], [38] and the development of vertical-cavity surface-emitting lasers (VCSEL's) (see, e.g., [39] and [40] for recent representative examples), were particularly important because they offered viable optoelectronic output devices for fabrication in large numbers in arrays. The optoelectronic device arrays and smart pixels stimulated significant work on array optical systems (see, e.g., [41]), leading to some large system demonstrations based on the SEED technology with, e.g.,

tens of thousands of light beams in a functioning digital system [42].

With the emergence of solder-bonding and related hybrid integration techniques for attaching optoelectronic input and output devices to silicon circuits [43]–[46], the smart pixel technologies evolved into dense optical interconnects to integrated circuits. With the hybrid integration of optoelectronic input and output devices directly over active silicon logic circuits [47], the distinction between smart pixels and optical interconnects is essentially lost since there is then no need to have the underlying silicon circuit formed as individual units associated with each input and output. The demonstration of the integration of thousands of quantum-well modulators and photodetectors to silicon logic chips shows that optics might be a feasible approach to dense interconnection to chips [48]. With large numbers of high-speed, low-power devices attached in this way, many benefits become possible for optical interconnects, including several that were not significantly exploited in long-distance communications.

A broad range of novel optical technologies has also been investigated to support possible optical interconnects to silicon chips. Systems have been demonstrated using relatively conventional lenses for imaging whole arrays of optical beams from one chip to another. Various microoptical techniques, such as lenslet arrays and diffractive optics, have also been investigated. Many of these techniques have been recently reviewed in [49]. References [28], [42], and [50]–[59] give examples of free-space systems investigated for optical interconnect applications. Waveguides on silicon chips have been investigated [60]–[62], though complete systems using these are not yet demonstrated.

In addition to interconnections involving both optical outputs and optical inputs, there has also been significant work in use of optics for clock distribution, where a centralized laser feeds an optical distribution network to drive optoelectronic clock receivers [63]–[83].

With this body of research work, it is now relatively well understood what many of the benefits of optics could be for interconnects, and there are technologies, at least in the laboratory, that can perform large numbers of high-speed optical interconnects between silicon chips. With this background, we will now look at some of the potential benefits that have emerged from this work. We can categorize the benefits of optics to some extent, though obviously this categorization is arbitrary, and some items could also appear in multiple categories.

B. Scaling of Interconnects

To start the discussion of benefits of optics, it is important to understand some of the physical limitations to the scaling of electrical interconnections. Since these limitations are significant, and optics largely avoids them, they are a relatively compelling reason for considering optics.

1) "Aspect Ratio" Limits to Electrical Interconnects: Conventional electrical lines all possess resistance. This resistance limits the rise time of signals. We can think of a simple electrical interconnect (or "simple line") as being one where we drive the line with a voltage step and

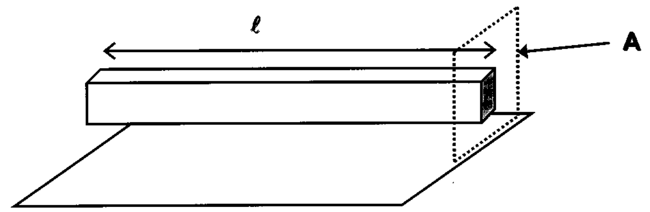

Fig. 1. Illustration of an electrical RC line. The area A is the effective total cross-section of the line and includes the space that must exist between lines.

detect the signal by measuring a simple voltage rise at the other end. In such a simple line, the rise time limits the rate at which we can send signals down the line. If we try to send bits of information too close together down the line, because of this finite rise (and fall) time, the bits will overlap (or, in communications terms, the "eye" will close in an eye diagram).

This rise-time scaling problem is easy to understand for the case of a "resistive-capacitive" (RC) line—one in which the capacitance of the line is charged through the bulk resistance of the line. A simple line is shown in Fig. 1. There are two conductors, here shown as a rectangular cross-section wire and a ground plane (the details of the line structure make no difference to this scaling argument). The wire has some effective cross-sectional area A; this area extends beyond the actual size of the wire, because we can only put adjacent wires so close together. The wire has a capacitance per unit length C_ℓ and a resistance per unit length R_ℓ. The total RC time constant of the wire is therefore $R_\ell C_\ell \ell^2$.

Suppose now we were to shrink the line in all three dimensions by some factor s, as might happen in a simple scaling down of a technology. The cross-sectional area would shrink by s^2, increasing the resistance per unit length accordingly to R_ℓ/s^2. The capacitance of the line per unit length does not change in such a shrinking—it depends only on the geometry of the line, not its size. (The capacitance of all well-designed lines is approximately a few picofarads per centimeter [4].) The length of the line has been shrunk to $s\ell$, and so the total RC time constant of the line is $(R_\ell/s^2)C_\ell(s\ell)^2 = R_\ell C_\ell \ell^2$—in other words, this shrinkage of the whole system by a factor makes no change in the RC time constant of the line. Almost all lines on-chip are of this RC-limited type. In general, lines with small cross-sectional areas tend to be RC lines. This leads to at least three problems: i) the transistors on a chip in general get faster as the technology dimension shrinks, so the wires are not keeping up with the transistors; ii) the chip itself likely is at least the same size in future smaller linewidth technologies, so we will tend to have actually longer delays to get from one side of the chip to the other; and iii) if we are running into a problem with interconnect delay in a given architecture, it is not solved by miniaturizing the system (we discuss this point in more detail below).

It is important to understand that this underlying scaling problem cannot be solved by redesigning the form of the line; the optimum design of line for minimum delay and optimum use of cross-sectional area is approximately one in which the separation of the conductors is comparable to the lateral size of the smaller conductor [4]. (Making the conductors closer

than this leads to too much capacitance. Making the separation larger than this leads to lines with too high a resistance for their cross-sectional area.)

We can generalize this argument [4]. This is straightforward for the case of the RC line. The total resistance R of a length ℓ of the line is $R \propto \ell/A$; total capacitance is simply proportional to length (i.e., $C \propto \ell$), independent of cross-sectional area (for a given geometry). Hence the RC time constant of the line scales $\propto \ell^2/A$. Since the number B of bits per second that can be sent down the line without excessive overlap of the voltage pulses corresponding to successive bits is $\propto 1/RC$, we find that $B \propto A/\ell^2$.

This limit on the bit-rate capacity B, of a simple line therefore depends only on the (architectural) "aspect ratio" of the line, by which we mean the ratio of the length l of the line to its cross-sectional dimension[1] (or more strictly, for a line with cross-sectional area A, the quantity $\sqrt{A}$). The existence of this limit has been understood independently by several authors [4], [2], [84] and is described in this "aspect ratio" form in [4].

Note that it would not matter whether we used one large cross-section line at a high clock speed or multiple smaller cross-section lines in parallel at lower clock speed; the total number of bits per second we could send this way would be the same (if all the lines were of the RC type), so this limit applies to the total cross-sectional area A of the wiring system.

If we use the conductivities of copper or aluminum at room temperature, assume lines that are well designed for best use of cross-sectional area, and require reasonably open "eyes" on eye diagrams, we obtain a limit for this kind of simple signaling on a simple line [4]

$$B \cong 10^{16} \frac{A}{\ell^2} \text{ bits/s } (RC \text{ lines}). \quad (1)$$

We can also analyze the situation for inductive-capacitive ("LC") lines. The rise times on such lines are limited by skin-effect resistance, which becomes worse at high frequencies. (Note that the transition from lines limited by bulk resistance to those limited by skin-effect resistance occurs at approximately the same frequency at which the line changes from being an RC line to being an LC line, so lines that are truly LC transmission lines are limited by skin-effect resistance, not bulk resistance.) The analysis of rise times on skin-effect lines is less intuitive than the RC case; by an accident of the algebra, however, the form of the scaling is the same as for the RC case, though with a slightly smaller prefactor, being [4]

$$B \cong 10^{15} \frac{A}{\ell^2} \text{ bits/s } (LC \text{ lines}). \quad (2)$$

This result is also counterintuitive in that we might expect LC lines to have better performance than RC lines. A key problem with LC lines is that the form of the rise time has a much longer "tail" than the RC case, leading to worse problems with overlapping of successive bits. This particular form of tail arises from the frequency dependence of the effective resistance that results from the skin effect. For high-speed systems, nearly all lines, other than on-chip lines, will be LC lines. Essentially, only lines on the chip are RC lines in such systems. (Long lines on high-performance chips are now crossing the boundary between RC and LC lines.) Note in particular that if we use simple signalling on the lines, there is a drop by as much as a factor of $\sim$10 in the number of bits per second we can get through a given cross-sectional area of line as we move from using RC to LC lines with increasing frequency.

Note that these "limits" are scale-invariant. The ratio A/ℓ^2 does not depend on the physical size of the system. This ratio is essentially an attribute of the architecture. Once we have drawn a block diagram of the architecture, we have essentially determined the ratio A/ℓ^2 with which we must work—it is a property of the number of wires that must come in and out of the walls of the "boxes" in our architecture, and how many "boxes" away we have to make the interconnections. Hence, for a given architecture, if we stay with the "simple" electrical wiring described here and try to increase the clock speed of the architecture, we will reach a limit on that speed at which we run into this "aspect ratio" limit. Then we cannot solve the problem either by miniaturizing the system or by making it bigger. Hence, we find wiring problems on chips, even though they are very small and have short wires.

We already routinely run into these aspect ratio limits when we have to make long connections between parts of the system. By these expressions, a 30-m-long unequalized coax line with cross-sectional area of 1 cm^2 could carry 100 Mb/s of simple digital data. [Equalizing (see below) might push this rate up to 10 Gb/s.] We are likely increasingly to run into the aspect ratio limit at any length scale as the aggregate data rates being communicated move into the terabit/second range.

One reason for mentioning the aspect ratio limit in connection with optical interconnections is that optical interconnections do not suffer from this aspect ratio limit. The physics of loss and signal distortion in optical interconnection is completely different, and there is essentially no distance-dependent optical loss or distortion over the scale of a machine (or even a computer room) with optical interconnects. (There are, however, losses in optical systems, though they are usually associated with components and connections, not with distance of propagation.) Hence, in an optical interconnect, if we can make optical connections to a chip, we can take those connections over essentially any distance without degradation. With electrical connections, even if we make an electrical connection off of a chip, to connect over any substantial distance is increasingly difficult, likely requiring wires much thicker than those used to connect to the chip itself. Hence with optics, we can contemplate kinds of architectures that are physically very difficult with electrical interconnects, or we can take existing architectures and continue to scale them to higher clock speeds without having to deal with the problem of the architectural aspect ratio.

Of course, there are good electronic engineering approaches that can get past these limits. For example, we can equalize the lines [2]–[4], we could use multilevel signaling methods, and we can put repeater amplifiers in the lines.

[1]The architectural aspect ratio defined here (a ratio of length to cross-sectional dimension) should not be confused with another meaning of aspect ratio that is the ratio of the height of the line to its width.

Equalized lines likely obey the same form of scaling [4], though with a somewhat larger prefactor, leading to

$$B \cong 10^{17} \frac{A}{\ell^2} \text{ bits/s (equalized lines)} \tag{3}$$

where the prefactor is limited by the signal level we can reliably detect in the receivers.

Sophisticated signaling methods, such as multilevel signaling, and coding of data on the line could be used, as is done in modems, in which case we could approach a Shannon limit for interconnection. This would require yet further complexity, however, and likely substantial latency in the signaling.

We can also divide a long line up into multiple short parts, joining the parts with repeater amplifiers. Such an approach is viable on chips because it is relatively easy to build the necessary amplifiers and connect them to the lines, and we will discuss this in the next section for on-chip interconnect. For off-chip interconnects, such repeatering would require coming on and off repeater chips, which makes it much less desirable.

The general conclusion here is that there is a relatively well-defined point at which we have to start bringing in further techniques if we stay with electrical interconnects, increasing the cost of electrical interconnection in terms of power, chip area, and complexity. Optics, at least in principle, can solve many of the problems seen in scaling electrical interconnects.

2) On-Chip Electrical Interconnects: The scaling problems discussed in the previous section make a strong argument for the use of optics for off-chip interconnects, where lengths are long and repeater amplifiers are awkward (if we had the optical technology at sufficiently low cost). These same scaling problems also exist on the chip, though there are some additional aspects (see also [26]). The potential reasons for using optics on the chip need a more detailed analysis than the off-chip case, and, with future scaling of the technology, require a more extensive analysis than current published ones (e.g., [26]). We will examine specifically only global interconnect lines, i.e., lines designed to communicate over sizes comparable to the chip size, to see if there is significant opportunity for optics at such size scales.

The scaling of microprocessors and their wiring structures has been discussed by several authors [85]–[90]. There has been considerable recent work, for example, by Deutsch *et al.* [91]–[96] and others [97], examining the effects of inductance, skin effect, and transmission-line effects in wiring on chips. The analysis of on-chip repeatered electrical interconnects was discussed extensively by Bakoglu [98] for the case of RC lines. A central conclusion of the recent work [91]–[97] is that inductance and transmission-line effects are becoming quite significant for long lines on CMOS chips, where previously such lines could be effectively modeled as RC lines. Other electrical options for long lines on chips include adding layers of external lines by solder bonding an interposer of copper lines and polymer dielectric onto the chip, with these lines also typically being LC lines [99].

The issue of the scaling of electrical lines including such inductive and transmission-line effects is particularly important for understanding whether optical interconnects have a significant role on the chip; we give a simplified scaling analysis in the Appendix and present some conclusions here.

We have deliberately chosen simple analytical models in this analysis. Such models will not give as accurate results simulations for any particular case, and the particular numerical results from the models are only approximate. The use of the models does, however, give a sense of how performance will scale, and why the system behaves as it does.

There are several important special aspects about on-chip interconnects.

i) Repeater amplifiers are viable on chips. They will consume significant power and will require vias through multiple metal levels to connect from the signal lines down to the amplifier circuits and back. They can, however, improve the bandwidth of longer interconnects.

ii) In addition to bandwidth, signal propagation velocity is also particularly important on chips, because chips are usually synchronous digital environments. Current architectures usually rely on keeping the data synchronous within a chip. Such synchronicity becomes difficult if we cannot propagate a signal from one side of the chip to another in less than a clock cycle, for example. (Chips are now at the point where asynchronous networks of on-chip units have to be considered because of the difficulty of maintaining global synchrony.)

iii) Lines on chips will almost always be RC, not LC, lines (though the effects of inductance in the RC lines is not always negligible). This point about the RC nature of lines on chips is not always appreciated, and is discussed in the Appendix. This difficulty of making useful LC lines significantly limits the propagation velocity of electrical signals on chip. (As mentioned above, long lines on chips are becoming mixed RLC lines.)

Figs. 2 and 3 show results of some simple modeling of repeatered and unrepeatered lines on chips. Fig. 2 shows bandwidth and Fig. 3 shows delay. This modeling is discussed in detail in the Appendix. The modeling considers 0.25- and 0.1-μm technology generations explicitly, as representative of known and future technologies, respectively.

From this analysis, we can deduce several conclusions about global on-chip interconnects.

a) Bandwidth:

i) Unrepeatered RC lines in 0.25-μm technology already do not have sufficient bandwidth for global on-chip interconnects at the clock rate (as is well known in practice).

ii) Repeatered lines do appear to have sufficient bandwidth at all clock frequencies, especially if one uses multiple thinner lines (though the delay would suffer in that case).

b) Delay:

iii) The effective signal propagation velocity in repeatered lines will likely not get substantially faster than that in current technology generations, and will be limited to a relatively small fraction of the velocity of light (e.g., 10–20%).

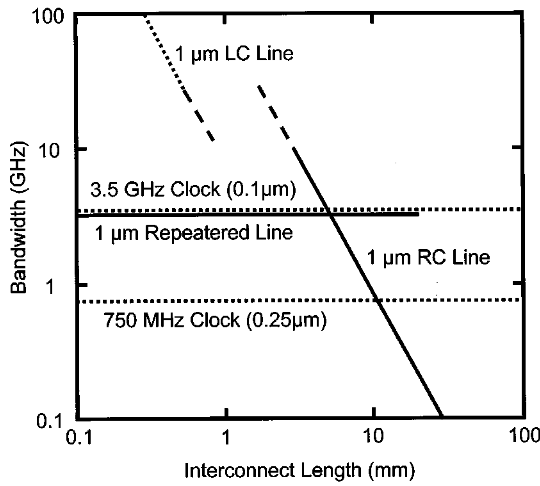

Fig. 2. Bandwidth of an upper level on-chip copper metal interconnect with a 1×1 μm cross-section, as a function of length. Both the case of an unrepeatered line (RC line) and a repeatered line (using 0.25-μm technology generation repeater amplifiers) are shown. Also shown are the on-chip clock frequencies for both the 0.25- and 0.1-μm technology generations. For sufficiently high frequencies, the behavior of the RC line would change toward that of an LC line, with shorter usable lengths, though pure LC behavior likely lies above the frequency range of interest.

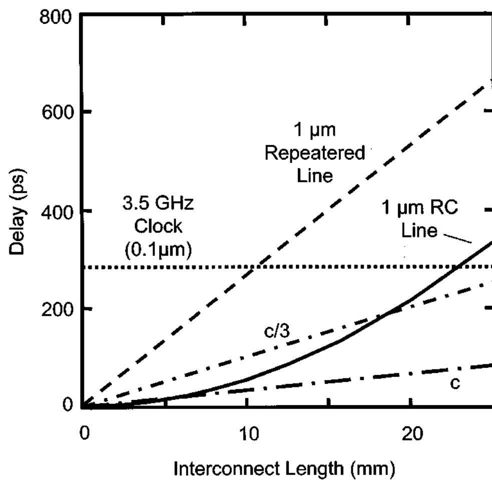

Fig. 3. Delay (rise time to 50% of final value) for copper repeatered and unrepeatered (RC) lines with 1×1 μm^2 cross-section. Also shown are the delay when propagating at the velocity of light in free space (c) and at $c/3$ and the delay corresponding to one clock cycle in the 0.1-μm generation. (The clock cycle for the 0.25-μm generation is $\sim$1.33 ns, which lies above the graphed region.)

iv) To carry clock-rate signals electrically across the chip at the local light velocity requires lines larger than the upper metal lines in current technologies.

v) To carry clock-rate signals electrically across the chip at the local light velocity in future technologies would require very large wires (e.g., as much as 4000 μm^2 cross-sectional area at 10 GHz) or equalization of on-chip interconnects.

vi) For 0.25-μm technology, the delay of either repeatered or RC global lines is still less than a clock cycle, but for future technologies both lines will have delays longer than the clock cycle.

vii) Communications at an effective velocity that is a substantial fraction of the free-space velocity of light (e.g., $>0.3\ c$) would allow global communications within less than one clock cycle even up to the 0.1-μm technology generation.

As mentioned before, the numbers deduced here are based on simple models. It is possible (and, indeed, likely) that good electrical design and use of thick metals in additional upper wiring levels can lead to electrical performance somewhat better than predicted here, but the underlying trends remain.

The conclusions about both the relative impracticality of LC lines and the relatively low limits to propagation velocity in global lines for future generations of chips may not to be widely appreciated in the community.

As far as optics is concerned, the relative difficulty of communicating across chip at a substantial fraction of the velocity of light with electrical interconnects may give an opportunity for optics in on-chip interconnects. The arguments here are, however, less clear than in the off-chip case. The optical technology would also have to be fast, with low latencies in the drivers and receiver circuits on the order of only a very few gate delays altogether. Optics could, however, be the only practical way to keep large chips synchronous as we advance to future technology generations.

3) Scaling of Optical Interconnect Driver and Receiver Circuits: As discussed above, as the transistors get faster, the wiring does not scale to keep pace. It is important to understand whether the necessary electronic driver and receiver circuits for optical interconnects could keep up with the clock rates of the logic transistors.

The design of the electronic circuits that should be used for the receiver amplifiers and the transmitter drivers for optical interconnects is still an open issue. This has been considered by various authors as part of their analyses [8]–[11], [19]–[21], [26], [29], [30], [47], [58], [101]–[104]. We discuss some of the challenges in receiver design in Section III-A1. It is certainly true that if we design receivers a) with large input photodetectors (e.g., with capacitances $\sim$1 pF) and b) for minimum detectable power, both of which are common criteria for telecommunications receivers, the receivers are far too large and consume too much power for use in dense optical interconnects. It is also true that the optical output devices, such as quantum-well modulators or VCSEL's, must operate at relatively low power so that the transmitter circuits do not consume too much power. With integration of relatively small photodetectors (e.g., solder-bonded photodiodes with areas $\sim 20 \times 20$ μm or less) and with use of received optical powers $\sim$1–10 μW, relatively low power dissipation and small size are possible in interconnect receiver circuits (e.g., in the range of a few milliwatts). At least with quantum-well modulators, and likely in the future with VCSEL's, total power dissipations in the range of a few milliwatts or less are also possible for transmitter circuits. (For example, a total electrical power dissipation on-chip of $<$6 mW was demonstrated with 0.8 μm CMOS at 375 Mb/s and 11 μW received optical power [47].) Given that optical interconnects may be possible with power dissipations low enough for dense interconnects, it is important to understand what would happen

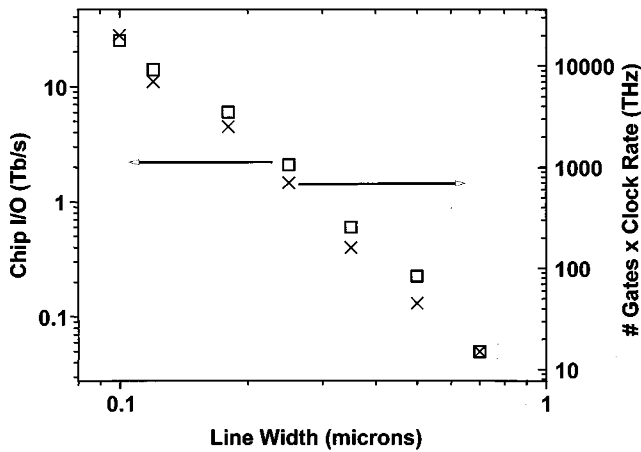

Fig. 4. The predicted I/O bandwidth and the "computational bandwidth" (product of the number of gates and the clock rate) of a CMOS chip with quantum-well modulator optical interconnects, scaled to future generations of silicon [21]. Despite the many different assumptions that go into these two calculations, the optical interconnects appear to be able to track the ability of the chip to perform logic operations.

in future generations of silicon technology. As the clock rates get faster, will the circuits for optical interconnects be able to keep up and achieve acceptable power dissipation?

This situation was analyzed [21] for the case of optical interconnects based on quantum-well modulators. Optical interconnect circuits do get better as the underlying transistors get smaller because the transistor transimpedances and other parameters of the CMOS transistors important for analog amplifiers improve with the scaling of the transistors. One of the key results is shown in Fig. 4. Here, the product of the number of gates on a chip and the chip clock rate is used to parameterize the increase in the ability of the chip to perform logic operations (though this is not intended to imply that all logic circuits are operated all the time). This product is compared with the calculated limit on optical interconnection on and off the chip. This result shows that optical interconnects may be able to keep pace with the ability of silicon chips to perform logic operations.

The calculation is based on predictions from SIA roadmaps, on scaling of silicon transistor performance, and on scaling of optoelectronic devices and integration. For example, the calculation for the 0.1-μm generation presumes 1 GHz interconnect rate (which may be conservative given more recent technological projections) and a total capacitance of the bonded photodetector of $\sim$20 fF. It estimates a received optical energy per bit of $\sim$10 fJ and a total power dissipation of transmitter and receiver circuits $\ll$1 mW. There are many assumptions that have to go into such a calculation, and one open question is whether the required density of receiver circuits can be built without suffering excessively from crosstalk. The calculation does, however, include power dissipation limits. One of the stronger limitations, if quantum-well modulators are used for the optical output devices, is that the power dissipation is limited by the receiver circuits, not the driver circuits. Other output devices, such as lasers or light-emitting diodes, could result in larger power dissipations limited by the driver circuits, though very-low-threshold lasers in principle could have similar power dissipations to modulator systems.

C. Clocking and Synchronization

1) Predictability of Timing: The delay on optical signal or clock paths is not strongly dependent on temperature, and signal or clock edges do not degrade substantially over the scale of a computer room. It is likely possible to retain absolute timing accuracy in the delivery of optical clock signals of $\sim$10–100 ps over a computer room (tens of meters) without any special technology. Ordinary optical fiber could be used. Optics cannot avoid the propagation delay from the velocity of light, and in fact propagation on optical fiber is somewhat slower than that on coaxial cable because of the refractive index of glass ($\sim$1.5), but the arrival time of optical signals is likely to be reliable and predictable. Such predictability could also likely apply to systems in which multiple beams have to be delivered to different places, such as in clock distribution. By contrast, the effective delay on electrical lines depends on temperature because the resistivity of copper depends on temperature. For example, copper's resistance changes by $\sim$40% over 100 °C. The rise time on electrical lines typically is proportional to the resistance of the line [this applies both to RC lines and LC lines (transmission lines)], and the signal delay on RC lines is proportional to the resistance. In addition, gate delay varies with fabrication and temperature.

There is also variability in delay from the driver and receiver circuits in both the electrical and optical cases. With the use of small optoelectronic devices integrated well with their electrical drive circuits, the optical circuits should have the same or fewer numbers of stages than their electrical counterparts, and hence comparable or better delay variation.

Additionally, because of the predictability of timing of optical signals, it could be physically possible to eliminate synchronizing circuits in interconnect links. We could, for example, imagine a system in which the clock is delivered optically in essentially perfect synchronism throughout the system. If an optical interconnect link from one part of the system to another is being driven by an optical modulator, the modulator itself can be read out by a short optical pulse synchronized to the optical clock; such a readout process resynchronizes the data to the timing of the short optical pulse. This resynchronization could actually remove any delay variation from the transmitter circuit. The removal of jitter and signal skew in an interconnect link has recently been demonstrated using an optical modulator and short pulse optical readout [100]. (We discuss this and other aspects of the use of short optical pulses in Section II-E7.) That data can then be sent over the optical link with a predictable arrival time, which could be set to be an integer number of clock cycles. There is therefore then no physical need for resynchronization at the receiving end—the data are arriving exactly synchronous with the local clock timing. This might be an important benefit for systems that need to process very large amounts of data, such as switching systems. Of course, system designers are used to including

synchronizing circuits and may wish to retain them to allow systems that are not necessarily globally synchronous, but optics offers the physical opportunity of eliminating them.

2) Reduction of Power and Area for Clock Distribution: In general, in optics, the power consumption does not depend on the distance of signal propagation. By contrast, at least on chips, the power required to distribute signals or clock will depend on the total length of the distribution lines. Optics could likely eliminate the need for high-power clock drive circuits on chip.

D. Design Simplification

One of the more subtle potential benefits of optical interconnects is that it may in the end make design of interconnect systems simpler. Optics avoids various issues that become increasingly troublesome in electrical interconnects as clock speeds and interconnect densities increase. In general, the design of an optical system itself (though not the optoelectronic devices) need not change as the clock speed of the system is increased, except that the delay in clock cycles will increase at higher clock rates.

1) Absence of Electromagnetic Wave Phenomena (Impedance Matching, Crosstalk, and Inductance Difficulties): Most of the difficulties of impedance matching and wave reflections can be avoided in optics, compared to the many problems encountered in electrical interconnect buses, for example. The impedance matching in optics is handled by antireflection coatings, which match impedance by zero-dissipation, "resonant" techniques that nonetheless have very broad effective bandwidths. Optical buses could use beam splitters that do not result in impedance discontinuities (in contrast with electrical buses where plugging in a new card can upset the impedance matching in the system). A common problem with electrical interconnects is reflection from capacitive loads. If the rise times of the signals have to be fast, then termination of the lines at one or both ends is generally necessary. Termination tends to increase power dissipation. Though capacitances of electrical circuits in optical interconnects are important, and may degrade performance, they do not result in reflection of the optical signal, and electrical termination can be avoided.

This ability of optoelectronic devices simultaneously to have matched impedance for wave absorption while still matching to the high (typically capacitive) impedance of the electronic devices is because of the phenomenon of quantum impedance conversion [11] that is intrinsic to all optoelectronic devices. Optical sources and detectors (and also some optical modulators such as quantum-well absorption modulators) are quantum devices. An electron of current in the device is associated with a photon emitted or absorbed by the device. The electrical impedance in the circuit connected to the device is irrelevant when considering the propagation and wave impedance matching of the emitted or absorbed photon. Thus such quantum devices separate the issues of the electrical circuit impedance and the propagating wave impedance. Effectively, this means that optoelectronic devices can be used as efficient and small impedance transformers to match the naturally high impedance of small electronic devices to the low impedance of wave propagation. Another way of expressing this advantage compared to the electrical case is to say that in the optical case we only have to provide enough charge to (dis)charge the capacitance of the photodetector, not the wiring.

Electrical connections to chips usually have to deal with the finite inductance of the connections to the chip, which means that changes in current can lead to substantial voltage errors; there is no analogous phenomenon in optics. For example, a short electrical wire (e.g., 5 mm) will have an inductance of $\sim$1 nH. Changing the current by $\sim$10 mA in $\sim$100 ps on such a wire results in an inductive voltage of $\sim$100 mV. Additionally, on medium and long lines, inductive crosstalk between lines can be important (see, e.g., [96]). There is also, of course, significant capacitive coupling and consequent crosstalk between adjacent lines on chips.

By contrast with electrical interconnects, optical interconnects do not generate or detect radio-frequency signals or interference. This is fundamentally because they do not measure classical voltages but rather count photons, and the only photons they can detect are those with sufficient photon energy (e.g., $\sim$1 eV).

2) Distance Independence of Performance of Optical Interconnects: As mentioned already, long optical interconnects perform just as well as short ones. The signals essentially do not degrade with distance, crosstalk does not increase with clock speed or with distance, and the power required to send signals optically is essentially independent of distance. Electrical interconnects between chips are usually significantly more difficult than those on-chip, for a variety of reasons including the capacitance of bond pads and wires, and the inductance of pins. Especially with free-space optics, interconnects between chips could be equal in performance to interconnects on-chip. Hence, with optics, we could avoid the need to design a hierarchy of different interconnects at different length scales, at least for off-chip interconnects. It is quite feasible in optics to make high-speed connections directly from a chip to another chip tens of meters away with drivers and receivers that are no different from those used for short optical interconnects.

An open issue in optical interconnects is how the optics itself would be made for dense interconnections between chips, and this issue will affect the practical distance dependence of optical interconnects. As mentioned in Section II-A, there are many possible approaches to the optics. We might expect some hierarchy of optical solutions. One approach might be to use rigid waveguides on the chip and on the backplane or board to which the chips were connected, followed by connection to flexible optical fibers for longer distances. If we only use such guided-wave channels, their numbers may be relatively limited. Consequently, it is more likely we would consider time-multiplexing the optical channels to make best use of them, which would result in more complex driving and receiving circuits, removing the simplicity of retaining similar driver and receiver circuits independent of distance. We could, however, imagine using free-space

optics with large arrays of beams operating perpendicular to the surface of the chip. In this case, the many beams can be handled at once with substantial economy of scale (e.g., with imaging optics). In this case, multiplexing the data is likely not necessary, and the optical interconnects could be run as simple digital lines, with transmitter and receiver circuitry no different for different distances. Free-space optics is only likely feasible within relatively rigid modules, but at least for signals sent within a module, it might be possible to keep the same simple driver and receiver circuits.

3) Frequency Independence of Optical Interconnects: The carrier frequency of optics is so high ($\sim 10^{15}$ Hz) that there is essentially no degradation or change in the propagation of the signals as the modulation frequency is increased, because the modulation frequency is negligible compared to the carrier frequency. (Only over long distances is the dispersion of propagation in optics important.) Hence the optics itself does not have to be redesigned as the clock speed is increased. The only change is that the delay in the optics will be a larger fraction of a clock cycle (or a larger number of clock cycles) as the clock frequency is increased.

E. Other Performance Benefits

Optics offers a number of other potential benefits to the performance of systems.

1) Architectural Advantages: There are various features of optical interconnect that could substantially alter the kinds of architectures that could be built.

Optics may allow larger synchronous zones in systems, not only on one chip (as mentioned above) but also possibly extending to multiple chips, even at gigahertz clock rates. (By synchronous zones, we mean ones in which the time delay is predictable, though not necessarily less than one clock cycle.) The limitations in the performance of electrical interconnects at high speeds are making it increasingly difficult to retain large synchronous zones in systems. Optics helps because the timing of the arrival and interconnection of signals is likely more predictable, and because effective signal velocities may be higher (compared to RC lines and/or repeatered lines).

As discussed above, an optical architecture is not constrained by "aspect ratio" limits—large numbers of long, high-speed wires are possible in optics. In particular, optics allows "fire-hose" architectures [105]—architectures in which very large amounts of information flow through the system. Such systems have problems that cannot generally be solved by the "system-on-a-chip" approach because the data may not be generated on the chip, or the sheer volume may rule out local memory. Examples include switching systems, special data-base systems where a given input must be compared with a very large amount of complex data, and bus-style architectures where large amounts of data flow over a bus between processors or between processors and memory.

Finally, it is worth noting that optics can be very good at regular interconnects of long connections—very large "perfect shuffles" could be made relatively easily with "free-space" optics.

2) Reducing Power Dissipation in Interconnects: Interconnect power dissipation may be reduced, especially for longer interconnects, because of "quantum impedance conversion" [11] (see Section II-D1). Other analyses of power dissipation have been performed with similar conclusions [9], [26], [29]. An additional source of saving in power dissipation may be that we can avoid building resynchronization circuits (e.g., phase-locked loops, buffers), see Section II-C1.

Taking full advantage of quantum impedance conversion to reduce power dissipation does require that the optoelectronic devices are well integrated with the electronic driver and receiver circuits (otherwise the total capacitance associated with the devices may be too large) and efficient optical output devices (in practice, either quantum-well modulators or very low threshold lasers).

Analyses of the relative power dissipation of optical and electrical interconnects typically can deduce a "break-even length" beyond which optical interconnects are energetically favorable. This length can vary from about 100 μm to tens of centimeters depending on the technological assumptions that are made.

There is an interesting debate on the fundamental limits of power dissipation in electrical and optical interconnects. Berglind *et al.* [30] have argued that electrical systems will have lower dissipation than optical ones if the system is limited by fundamental signal-to-noise constraints and the power dissipation of the receiver is ignored, and comparable dissipation if receiver power dissipation is included. It is not clear, however, that these arguments apply to practical interconnects [106].[2]

3) Voltage Isolation: Optical interconnects intrinsically provide voltage isolation between different parts of the system. This is another consequence of the quantum nature of optical sourcing and detection. Optical detectors essentially count photons, not measure classical voltage, and provide perfect voltage isolation as a result. This benefit of optics is already exploited in optical isolators (combinations of light-emitting diodes and photodetectors). This could

[2]Berglind *et al.*'s [30] fundamental argument is based on the practical reality that optical systems must work with relatively large quanta of energy that are much larger than $k_B T$ (where k_B is Boltzmann's constant), whereas electrical systems can detect a bit of information with $\sim k_B T$ of received energy. This fundamental argument is largely correct, though it is interesting to note that it is also possible (though apparently *highly* impractical) to transmit information with $k_B T$ of energy per bit optically, even when $k_B T$ is much less than the photon energy. The trick to communicating multiple bits with one photon is to put the photon into only one of a large number N of possible time slots (or spatial or wavelength channels, or some combination of all of these). When we measure which slot the photon is in, we recover multiple bits of information ($\approx \log_2 N$). The limitation to the number of slots we can use is given by the random thermal excitation of other photons into the slots. In [106], calculations are performed for 1.5-μm wavelength photons at room temperature, which gives a limit of $\sim$42 bits per photon and requires the use of $N = 2^{42}$ time slots. Such a system also uses $\sim k_B T$ energy per bit. The argument of comparable dissipation for electrical and optical systems if receiver power dissipation is included is correct if we can send pulses down 50-Ω lines without loss; it is not correct if we consider that we send signals in electrical systems, at least over short distances, by fully charging the line, in which case we need to compare the line capacitance to the photodetector capacitance, and it is also not correct if we consider sending signals over longer distances where loss in the electrical line is important. It is also not likely in a digital environment that we would work with very low signal voltages because of all of the digital noise. Hence it is not clear if there is a substantial region in which Berglind *et al.*'s analysis [30] is correct for practical interconnects.

be a significant and important reason for adopting optical interconnects, essentially making systems easier to design because varying relative voltage levels and effects such as "ground bounce," which can be caused by inductive or resistive transient voltage drops on lines, can be avoided.

Voltage isolation may become even more important in future generations of electrical chips because supply voltages are reducing, thereby reducing tolerance to voltage variations, and supply currents are increasing, thereby increasing both resistive variations in dc voltage levels and ground bounce effects due to line inductance.

4) Density of Interconnects: There is little question that electrical interconnects offer the highest densities for very short interconnects, such as those found locally on-chip. For the foreseeable future, optics does not offer any competition to such interconnects. The situation changes as we look at off-chip interconnects, and possibly at longer distances on-chip.

Optics can offer large overall densities of interconnects, especially for longer distances (i.e., cross-chip, off-chip, and interchip). For example, an experimental chip has been demonstrated that has over 4000 optical I/O's in a 7 $\times$ 7 mm area [107]. Electrical chips can have large numbers of pins (though usually not at densities as high as this example), but there will be multiple pins for each interconnect because of the necessity of having signal grounds, and in addition it will likely be necessary to use multiple pins for power and ground because of the pin inductance.

The limit on the density of optical interconnects to chips is likely power dissipation in the receiver and transmitter circuits [21], at least for the case of quantum-well modulators as output devices. The receiver dissipation results primarily because the front end is a small signal amplifier, and thus requires steady bias current to keep the devices in the middle of their amplifier region. Nonetheless, it does appear that power dissipations in the low milliwatt or submilliwatt range may be possible for these circuits [21], which would allow thousands on a chip without special heat sinking.

There is also the simple geometrical effect in optics that we can exploit the third spatial dimension relatively effectively for interconnects. In a conventional electrical interconnect, we may have two-dimensional parallel interconnection from the chip to the circuit board by use of area solder bonding, for example, but we then have to "squeeze" the information back into "one-dimensional" propagation in a plane. In optics, when we connect in parallel in two-dimensional arrays perpendicular to the surface of the chip, we can continue to exploit the third dimension relatively effectively, for example, through the use of imaging lenses or even fiber bundles. Only if we can effectively implement three-dimensional electrical interconnects or 3-D VLSI can electrical interconnects have a similar advantage. Note too that we often use the third dimension in electrical interconnect just to allow crossing wires (e.g., by going up to a higher wiring layer). In free-space optics (and even in some waveguide optics), we can instead exploit the fact that light beams can pass through one another without degradation.

It is worth emphasizing that, in optics, it may not be necessary to go through a hierarchy of interconnects in order to connect to distant points. Every one of the interconnects off of a chip could be a long-distance interconnect if required, without further repeaters or drivers. By contrast, in electrical systems, if we are considering thousands of interconnects off of a chip, it is likely to be difficult or impossible for many of the interconnects that come off a chip to be propagated very far.

Though optical fibers may be subject to similar limitations in density as short-distance electrical interconnects, free-space (e.g., imaging) optics allows a substantial economy of scale in handling regular interconnects of very large numbers of inputs and outputs with its ability to handle thousands of light beams with only a few optical elements, and could retain very high interconnect densities.

We have also argued above that as clock frequencies increase, it becomes increasingly difficult to send signals across chip with low signal delays. If we try to send signals across the chip at near to light velocities, the resulting LC transmission lines will have to be relatively large (e.g., $\sim$63 $\times$ 63 μm^2 cross-section for 10-GHz signals). There may even therefore be an argument for the use of optical interconnects to enable relatively dense global connections on the chip. Waveguides can be contemplated on chip with cross-sectional dimensions of the order of several micrometers [60]–[62] (though devices to drive optical signals efficiently into such waveguides are still problematic), and free-space devices with areas $<$10 $\times$ 10 μm appear quite feasible. Though a 10 $\times$ 10 μm optical device is large compared to the micrometer widths of a wire, it is small compared to the total area (length $\times$ width) of a global wire on a chip.

5) Fabrication Benefits: If we decide to use optics for longer distance interconnects on-chip, then we may eliminate the need to develop the technology for yet more wiring layers [33], though of course we would have to develop the optical technology and the associated electronic receiver and transmitter technology instead.

6) Testing: With optoelectronic input and output devices on chip, it is possible that the chip could be tested in a noncontact optical test set. Because optics can communicate high-speed signals without degradation over substantial distances, the test equipment can be remote from the testing, and high-speed electrical probes can be avoided. The numbers of optical beams could be large, and the parallelism possible with optics might be able to accelerate the testing of chips. For example, there is no basic reason why in an optical system we could not test thousands of points on chip at once. In such a test set, it would only be necessary to make the electrical power supply and low-speed control connections by conventional electrical probing, while providing all other high-speed and large-number test abilities optically. Note too that the optics need not physically touch the surface of the chip in order to do the testing.

7) Benefits of Short Optical Pulses: One radical opportunity in optics is the use of short optical pulses to power interconnects. In optics, it is relatively straightforward to make trains of very short light pulses (e.g., 10 ps–100 fs) through the technique of laser mode-locking. (Even shorter pulses can be made with somewhat more effort.) There is a variety of ways of performing such mode-locking, and the repetition rate of the pulses can be set from tens of megahertz up to

hundreds of gigahertz (see, e.g., [108]). Such sources would require some development for use in computing systems, but they are routine in the research laboratory.

Hence we could imagine using a centralized mode-locked laser as a clock source and/or a synchronized optical power source for driving modulator-based interconnects. The use of such a mode-locked source raises the possibility not normally available in electrical systems, which is that we have pulses that are much shorter than the rising or falling edges available from the electrical circuits. This has several possible benefits.

Perhaps the most obvious benefit of such short pulses is in the precise delivery of clock signals. Note that with optics, we can deliver such short pulses even over long distances without substantial degradation. We can certainly imagine a centralized mode-locked clock source distributed quite precisely over an entire system. Note too that the signal delivered has very fast rising edges, which means that the clock phase could be very well defined.

The use of short pulses may also allow improved receiver performance. With a short pulse input, we can obtain the impulse response of a receiver, which, for a given optical energy, will give a larger transient peak voltage output than we would obtain if we drove a receiver with the usual relatively slow rising edge [109].

If we use short optical pulses to read out optical modulators, we can intrinsically resynchronize the data. We could imagine, for example, that we have optical modulators connected to each of several different electrical outputs from system. Because of the usual skew that exists in the electrical system, the different electrical signals would have slightly different timings. If, however, we wait until all of the electrical signals are valid, and then read out all of the modulators with the synchronized set of short optical pulses, we will have removed the skew from the signals and will have consequently resynchronized all of the data without the need for electrical buffer circuits. Such resynchronization has recently been demonstrated [100]. As mentioned in Section II-C1, such resynchronization might allow large synchronous systems.

Short optical pulses also offer a radical method for making wavelength-division multiplexed interconnects. A short pulse (for example, 100 fs) intrinsically possesses a large optical bandwidth (for example, 10 nm wavelength range). By dispersing the different wavelengths over different modulators, we can modulate each channel of information in a different wavelength band, put all the wavelength bands back together into a single light beam by passing them back through the dispersing system, and hence send a multiple channel interconnect over one optical beam or optical fiber to another chip (on the other chip, we would have a similar dispersing system that would separate the wavelengths out again onto different detectors) [110].

III. Challenges for Optical Interconnect

Though optics is very attractive for interconnects for a variety of reasons given above, there are many practical challenges.

Stated simply, the main problems for optical interconnect are that:

i) the technology is still immature and costly;
ii) the systems that could take most advantage of optics likely have architectures different from the current architectures that are optimized around the strengths and weaknesses of electrical interconnects.

There is also a significant issue that the problems and benefits of optical interconnects are often substantially misperceived by those not involved in recent research work. This kind of situation is a common one for a new technology, but it represents a significant barrier for the successful introduction of optical interconnects. We also briefly discuss some of the misconceptions below.

A. *Specific Technical Challenges*

There are many technical challenges in implementing dense optical interconnects to silicon CMOS chips. These include circuit issues, especially for receiver circuits, necessary evolutionary improvements in optoelectronic devices, integration technologies, and, especially, the development of appropriate optical technology to allow low-cost optical modules.

1) Receiver Circuits and Low Capacitance Integration of Photodetectors: Integration of detectors is very important for receiver performance because low input capacitance is essential if receiver circuits are to be kept small and their power dissipation is not to be too large. Receiver power dissipation may well turn out to be the largest power dissipation in optical interconnects [analyses (see, e.g., [21] and [25]) vary in their conclusions here, depending on the specific assumptions made about optoelectronic devices, though receiver and transmitter dissipations are expected by most authors to be roughly of comparable magnitude].

Large detectors and/or large associated capacitances mean that more sensitive amplifiers have to be used. Large input capacitance can also mean that the Johnson/Nyquist noise of the transistor channel is more troublesome and can tend to require the use of larger transistors with associated larger power dissipation in the receiver. A standard optimization for telecommunications receivers [111], for example, is to set the transistor capacitance approximately equal to the total detector capacitance, an optimization that approximately minimizes the effect of Johnson/Nyquist noise, but that leads to very large input transistors with high power dissipations. More sensitive amplifiers require more stages, increasing latency and power dissipation. Very sensitive amplifiers will also be more sensitive to electrical noise, such as power-supply line noise and other digital noise in the system.

A better approach for optical interconnect receivers is to make the physical capacitance of the photodetector and its connection to the receiver circuit as small as possible. Small capacitance leads to larger voltage swings for a given optical energy, which leads to better noise immunity and fewer gain stages. Smaller capacitance also allows the use of small, low-power-dissipation transistors in the input stage. There is no particular need to make receivers with the kinds of sensitivities used for telecommunications [e.g., $\sim$150 photons (20 aJ) per bit]. Received energies of $\sim$1–10 fJ are more likely a better optimum choice for interconnect systems.

A design goal should be to keep the receiver power dissipation in the low milliwatt range or below so as to keep the overall power dissipation of the chip sufficiently low when using large numbers of optical interconnects. As mentioned above, systems with 3.5 and 2.0 mW for receiver and transmitter power dissipations, respectively, have been demonstrated at 375 Mb/s [47], with circuits that can be surprisingly small (e.g., $17 \times 18\ \mu m^2$ area), and require relatively little optical received power (11.5 μW per channel) or received energy per bit ($\sim$30 fJ/bit), so milliwatt dissipation seems an attainable goal with future CMOS. Even with the $\sim$50 fF capacitance of current solder-bonded p-i-n diodes [47], received optical energies $\sim$10 fJ can allow input voltage swings $>$100 mV, which may be sufficient for noise immunity, and allow receivers with small transistors and only a few (1–3) stages of amplification to recover a logic level. For other recent examples of optical interconnect receiver design, see, e.g., [100]–[103].

A speculative possibility for optical interconnect receivers is to run "receiverless"—that is, have the detector capacitance so low that the input optical signal directly drives a logic-level voltage swing in the input photodetector. Such an approach has many advantages, though will require good integration of the photodetector, and small photodetectors. For example, with $\sim$10 fJ optical energy per bit and an input detector capacitance of $\sim$3 fF, voltage swings of 1 V or larger could readily be attained directly in the photodetector. Three femtofarads is not an unphysical level of capacitance to be achieved in a small photodetector (this is about the capacitance of a $5 \times 5\ \mu m^2$ p-i-n III–V photodiode or a somewhat larger M-S-M photodetector), though monolithic integration would be required to keep the stray capacitance at comparable levels. Such "receiverless" systems would have very low latency.

In all concepts for dense optical interconnects directly to silicon, the practical issues of receiver crosstalk and noise immunity have not yet been addressed sufficiently.

2) Evolutionary Improvement in Optoelectronic Devices: Quantum-well modulators and VCSEL's are strong candidates for viable output devices for dense optical interconnects to silicon. No major breakthrough is required for use of these devices, at least with hybrid integration. In the case of quantum-well modulators, the same devices can also be used effectively as input photodetectors, giving a potentially viable complete solution. In the case of VCSEL's, different photodetectors are required, though there is no basic problem in making efficient photodetectors in III–V technology; recent work has also shown how lasers and photodetectors can be fabricated using the same structure, for example, [112].

Light-emitting diodes (LED's) are advocated by some as output devices. They have the advantage that they may be relatively easier to make than VCSEL's and might avoid some of the problems of lasers, such as mode and polarization stability. They have disadvantages of limited speed of response and overall optical efficiency, which we discuss below. Silicon-based output devices (light emitters and modulators) do not appear viable at the present time. Below, we summarize some of the areas that will need evolutionary improvement in optoelectronics.

c) Quantum-well modulators: Quantum-well modulators have so far been the devices most extensively used in demonstrating actual dense interconnects to and from silicon CMOS chips. They have successfully been made in large arrays that have also been solder bonded to the circuits. Their performance has been good enough to allow demonstration of large optically interconnected laboratory systems (see, e.g., [42], [48], [58], and [113]). When integrated with CMOS, the quantum-well technology is sometimes described as CMOS-SEED or optoelectronic VLSI.

Modulators require that an external beam be brought onto the modulator. This requirement can be considered a disadvantage or an advantage depending upon the overall requirements of the system. Obviously, to bring in external beams requires more optics to generate beam arrays and to handle the separation of incident and reflected light beams. (Most systems with modulators operate in reflection, so that it is not necessary to make the chip and its mounting transparent.) If, however, the optics is set up to handle these beams, it is only necessary to generate and control one master laser beam, for example, in wavelength, amplitude, and mode quality. Such a master beam can be split up by diffractive optics into the necessary arrays of equal beams. Additionally, the use of a single master laser allows centralized clocking of the entire system, and the use of modulators, as described above, allows the retiming of signals, especially if the master laser operates with relatively short optical pulses. This ability to synchronize the system may turn out to be a significant advantage for modulator systems. In addition, the optics required to power modulator arrays is essentially similar to the optics required for clock distribution to a chip. Modulators avoid many of the problems of mode quality, wavelength stability, turn-on delay, and, arguably, power dissipation that remain significant issues for VCSEL's at least with present VCSEL technology.

To be compatible with the lower voltages of future generations of silicon, and to have the larger contrast ratios and wider wavelength and temperature tolerances that may be required in practical systems, it will be necessary to investigate concepts such as stacked [114] or interleaved [115] diode structures or modulators with resonators (see, e.g., [116]) that can likely address these requirements.

d) Vertical-cavity surface-emitting lasers: VCSEL's have made substantial progress in recent years, especially with the advent of oxide-confined structures that promise lower threshold currents [117]. Various improvements are desirable for practical use in dense interconnects on or between chips. Compared to quantum-well modulators, VCSEL's have also seen less use yet in systems with large arrays operating with silicon circuits, and so the practical issues for use in large dense interconnect systems are less well understood. Possible issues include the following.

i) *Threshold currents*. For large-scale dense use, it is likely necessary to achieve threshold currents in the range of tens of microamps. Such thresholds would allow output powers of $\sim$100 μW while still avoiding

problems of "turn-on" delay [118]. Turn-on delay is a phenomenon in which there is a delay in the emission of light from the laser that depends on the previous data pattern. It can be avoided by careful biasing of the VCSEL, though this requires more complex drive circuitry. The other method of avoiding the timing variability of VCSEL turn-on is to arrange always to drive the VCSEL substantially above threshold. Such low thresholds likely require the use of "oxide-confined" VCSEL's, and large arrays will be required with predictable wavelengths and thresholds. Dense arrays of VCSEL's with higher current densities can also run into thermal problems.

ii) *Mode and polarization control*. VCSEL's are prone to having different spatial optical modes at different currents. This is highly undesirable for use in interconnects, since it can cause variation in the detected power, generating spurious signals. For use in dense systems, the VCSEL should retain single-spatial-mode operation likely up to ten times the threshold. This may be possible with oxide-confined VCSEL's, though work remains to be done to make this a predictable property. Some optical systems can be sensitive to optical polarization, and it is desirable to control this. Control of VCSEL polarization properties is still a subject of research (see, e.g., [119] and [120]).

iii) *Wavelength control*. VCSEL's run at specific wavelengths set by the physical dimensions of the laser cavity. It is difficult to control this wavelength to better than a few percent, which may not be good enough to allow the use of diffractive optics in the optical system. The wavelength of operation also varies with temperature because of the change of refractive index with temperature in the cavity. (The properties of diffractive optics usually change in proportion to wavelength.)

iv) *Beam size.* Low-threshold VCSEL's necessarily have very small output beams (e.g., micrometers), which can create a difficulty in lining them up with optics for predictable beam directions.

v) *Power-supply voltage*. VCSEL's may have difficulty reaching the very low power supply voltages (e.g., <1 V) that are expected in future silicon CMOS. Since they will require voltages comparable to or larger than the bandgap energy in electron-volts to forward bias the devices, separate bias supplies may be required, which are generally considered undesirable.

e) Light-emitting diodes: Light-emitting diodes are still possible candidates for output devices in optical interconnects, but they suffer from at least two substantial problems that would have to be weighed carefully in any serious application in dense interconnects.

One major problem is that the speed of response of light-emitting diodes tends to be limited by carrier recombination times. To make efficient diodes, these times should not be artificially shortened, and this makes the idea of devices running at gigahertz speeds difficult though not impossible. One approach is simply to drive the devices with relatively high current densities to create relatively high carrier densities so that the spontaneous emission lifetime is relatively short.

A second major difficulty with light-emitting diodes is that it can be very difficult or impossible to collect all of the light from such an incoherent emitter and focus it efficiently onto a small detector. Light-emitting diodes generally emit over a very broad range of angles, and there are fundamental physical problems (the second law of thermodynamics, in the form of the constant brightness theorem) that can prevent such light being focused efficiently to a small spot. Hence the use of light-emitting diodes could lead to excessive power dissipation in the system. It is not clear how these problems with light-emitting diodes can be solved effectively, though there are serious attempts to deal with the efficiency and angular spread issues [121], [122] through the use of microcavities. The use of LED's is likely viable at least for moderate numbers of interconnects (e.g., tens) and at speeds up to a few gigahertz.

f) Silicon-based optoelectronic devices: Photodetectors made in silicon are in principle usable, though, at least when fabricated in the silicon CMOS process, they are not ideally suited to the wavelengths where the currently viable optical output devices like to operate (in the near infrared from 850 to 980 nm). One key problem is that the absorption length is quite large ($\sim$7 μm at 850 nm), much larger than the typical depletion thicknesses in the CMOS process. This length leads to two difficulties: i) the efficiency of the detector is low and ii) carriers created deep in the structure can diffuse over relatively long times into the depletion region of the detector, where they give rise to long tails in the detector time response. There has recently been some innovative work [102] in trying to exploit silicon CMOS for detection, however, that can avoid at least the problem of the long tails.

Using silicon-based materials systems for light emission for optical interconnects still would require major breakthroughs. Though several schemes have been investigated (see, e.g., [123]–[130], and a review of some of these techniques [131]), these techniques all apparently still have very low efficiencies of light emission (e.g., 10^{-4}–10^{-6}) even as incoherent emitters. Silicon can be used to make optical modulators [132], but these devices tend to be large because the optical effect used is relatively weak, and may also be too slow and have too large a required current drive for practical dense interconnections; they also only work in a waveguide configuration. Hence silicon optoelectronic devices, other than possibly photodetectors, do not appear viable for dense optical interconnect at the present time.

g) Centralized clocked lasers: As mentioned above, centralized lasers for use in optical clock distribution or for powering modulator-based interconnect systems have several advantages, including centralized wavelength and mode control, and mode-locking to produce short pulses. There does not appear to be any basic problem to prevent the development of practical mode-locked laser sources, such as diode-pumped solid-state mode-locked lasers [133], or even directly mode-locked diode lasers (see, e.g., [134]). There would be significant development effort required, however.

3) Absence of Appropriate Practical Optomechanical Technology: Undoubtedly, one of the more serious issues in implementing dense optical interconnects is the absence of low-cost practical optomechanical technologies that would allow us to make inexpensive optics modules to connect chips. It is relatively obvious that the kinds of optical technologies familiar to most people, such as camera lenses, are not suitable for optics to connect to chips. Such optics is too bulky for any serious system.

As mentioned above, in addition to the optical fibers with which most of us are now familiar, there are several other microoptical technologies that are feasible. Examples include lenslet arrays, both refractive and diffractive, that allow very large numbers of very small lenses to be made relatively inexpensively and with lithographic precision, diffractive optical elements that allow relatively complex interconnect and beam array generation patterns to be made with lithographic techniques, gradient index optics, imaging fiber bundles, fiber arrays, various waveguide technologies (including waveguides on silicon), and micromachining and plastic molding techniques that could allow relatively complex microoptical assemblies. There is not space here to discuss all of these techniques; many of these have been recently reviewed in [49].

For free-space systems (that is, those relying on imaging optics techniques), there does not appear to be any need for a physical breakthrough to implement relatively practical optical systems. The need is primarily in developing practical low-cost schemes from the various available technological ideas.

For waveguides on silicon chips, there still appear to be some basic technological and fundamental challenges. There are promising ways [60], [61] to make waveguides using processes that are compatible with silicon processing, though there are still some issues with waveguide loss, and it is not clear what optical emitter or modulator devices could be used if dense interconnects are required. As discussed above, there is little current prospect of small modulators or coherent emitters in a silicon-based technology, so overall efficiency remains an issue. It is also not clear how to couple waveguided light on the chip in a simple and efficient way to optics that is off the chip. It is encouraging that at least the waveguides themselves appear viable in a silicon process, and these may be interesting at least for optical clock distribution, for example, where on-chip sources are not required.

4) Integration Technologies: It is clearly necessary to be able to integrate the optoelectronic devices with the silicon integrated circuits. For the foreseeable future, it appears that we will need to use III–V optoelectronic devices at least for optical outputs in dense interconnects. It is currently very difficult to integrate lasers monolithically on silicon substrates. The dislocations that are formed in the lattice-mismatched interface apparently tend to propagate and degrade the laser performance over time. This does not appear to be a major problem for the case of quantum-well modulators, which have been successfully integrated with silicon substrates without apparently lifetime degradation [135]. There is, nonetheless, continuing encouraging work toward successful monolithic growth of III–V light-emitting devices on silicon substrates (see, e.g., [136]).

Even with such successful growth on silicon substrates, however, significant issues remain for growth on silicon CMOS circuits. For example, it is unlikely that a manufacturer would want to allow Ga in a silicon production line, since it tends to make the silicon oxides conducting. The standard metallization processes for III–V devices are more usually gold-based, which has compatibility difficulties with the aluminum-based metallizations of silicon CMOS. The present successful growth of modulator devices on silicon substrates [135] uses substrates slightly tilted from the usual on-axis direction. Growth of III–V materials on silicon substrates also typically requires a cleaning process that is likely at too high a temperature for the metallization on the chip. It is likely that these various compatibility issues could be resolved, at least for modulator and photodetector devices, but having to resolve such issues just to introduce optical interconnects would impose a significant barrier; there would be significant resistance to changing basic parts of the CMOS processing just to accommodate optics.

A more viable approach, especially for introduction of optics, is to use hybrid integration technologies such as solder-bonding. Such techniques require no modification of the basic CMOS process, using instead only a few simple additional metallizations that can be done on the finished wafer. This allows the separation of the growth and processing of the optoelectronic devices from that of the silicon circuits, and allows a great deal of flexibility in the types of optoelectronic devices used. There are several variants of such techniques, and these have been reviewed recently [45], [46], [137]. These techniques will require continued work, but they appear to offer a practical solution to integration for dense optical interconnects to chips.

B. Misperceptions

There are also some perceived problems with optics that are not in fact basic problems. These misperceptions are a significant problem for the introduction of optics.

1) "The wavelength of light is large and hence optics will never compete with electronics": It is true that optical devices themselves are not likely to be made as small as electronic devices for precisely the reason that the wavelength of light used (and likely to be used for the foreseeable future) is significantly larger than the size of the smallest fabricable electronic devices and interconnect wiring widths. However, for interconnects, the same logic does not apply. The true comparison is the total size of the interconnect circuit (and also possibly bonding pads) in the electrical case compared to the total size of the optical interconnect circuit and optical "output pad." Because long electrical connections require substantial driver circuits, the optics can "win" here to the extent that its driver and receiver circuits can be smaller, and its devices can be smaller than electrical bonding pads.

2) "Optical interconnects have substantial latency" and "the conversion from optics to electronics or vice versa is inefficient and requires too much power, area,

and time": Both of these statements can be valid when discussing the current generation of long-distance optical communications technology. In such applications, there is no particular requirement for efficiency or low latency. It is, however, important to emphasize that the conversion between optics and electronics can be extremely efficient. Most semiconductor photodetectors, for example, operate at essentially 100% quantum efficiency, and modern quantum-well modulators and very-low-threshold lasers are also highly efficient. The key technological requirement to keeping the overall power efficiency high is to perform very effective integration of small devices with the electronic circuitry. Examples quoted above (for example, [47]) show that efficient circuits are possible for complete optical interconnects. There is also no reason why well-integrated optical interconnects should have high latency. The circuit latency would essentially reduce to the delay through the total number of stages involved in driving and receiving the optical interconnect. Since optical interconnects likely require less drive than long electrical interconnects, fewer stages should be required at the driving end. With good integration, it should be possible to keep the receiver circuits to somewhere between one and three stages of amplification. Hence the circuit latency in an optical interconnect should only be a small number of gate delays. Note also that the same circuit latency would apply to long optical interconnects. The propagation latency of optical interconnects could conceivably be slightly longer than that of good coaxial lines, but the latency in propagating over shorter distances may well be much less than the effective propagation latency on repeatered electrical lines.

IV. Conclusions

The use of optics to make connections within and between chips could solve many of the problems experienced in current electrical systems. Many of the physical reasons for the use of optics are well understood and indicate many potential quantitative and qualitative benefits. Though there are, and will continue to be, electrical solutions that stretch the capabilities of electrical interconnects, optics is arguably the only physical solution available to solve the underlying problems of interconnects, and has the potential to continue to scale with future generations of silicon integrated circuits. Optics may also solve the growing problems of system synchronization, allowing, for example, individual silicon chips to remain synchronous domains even as the chips grow. These synchronization advantages apply both to optical clock distribution and to optical interconnects themselves. Optics may reduce on-chip power dissipation in clock distribution, global on-chip interconnects, and off-chip interconnects. Optics also may relieve a broad range of design problems; for example, optics itself does not have to be redesigned as system speed is increased, optical interconnects provide voltage isolation, and problems of wave reflection, impedance matching, and pin inductance are essentially absent.

The basic devices and laboratory demonstrations of the key elements of technology exist for dense optical interconnects to silicon IC's. No physical breakthrough is required to implement optical interconnects (though there are possibilities for revolutionary advances). Substantial technological work remains, however.

a) Optoelectronic devices require continued development to meet the yield, tolerance, and drive voltage requirements for practical systems with future generations of silicon CMOS.

b) Work will be required in the interface circuits between optics and electronics. Though there appears to be no current fundamental difficulty in making such circuits in CMOS, research is needed in circuits that i) avoid issues such as crosstalk and susceptibility to digital noise, ii) have appropriately low power dissipation and latency, and iii) are tolerant to process variations.

c) The technology for integrating optoelectronics with silicon integrated circuits is still at an early stage, though there have been key demonstrations of substantial working integrations. Likely first introductions of optical interconnect to chips will use hybrid approaches, such as solder bonding; such hybrid approaches require no modifications to the current process for fabricating silicon integrated circuits except to add processes to fabricated silicon integrated circuit wafers. Solder bump hybrid integration also simplifies design since the optoelectronics is then attached only to the top-level metal, allowing complete freedom of placement of electronic circuits underneath. Longer term approaches will likely examine monolithic integration, though such integration is still at a basic research level.

d) Novel approaches will be required for the optics and the mechanics for such optical interconnections, though there are several opportunities for appropriate optomechanical technologies.

e) It will be important to research the systems and architectural benefits of optics for interconnects. Optics can likely enable kinds of architectures that are not well suited to electrical interconnect systems (e.g., architectures with many long connections, architectures with large "aspect ratios," architectures requiring synchronous operation over large domains), and can likely also allow continued use of current architectures that otherwise would have to be abandoned in the future because of the limitations of wired interconnects. There are also significant opportunities for optics in the somewhat simpler application of clock distribution.

f) Undoubtedly, the issue of cost is a major one for optical interconnects. Nearly all the work on optical interconnects discussed in this article is laboratory research on possibilities. It largely does not enable even rough estimates of cost. Certainly, none of the optical research discussed here shows interconnects that cost less than current electrical interconnects used in any high-volume product. To some extent, it is not fair to

ask for cost estimates based on such "one-off" laboratory work. Cost depends crucially on volume. If a laboratory researcher was asked to make a compact disk player without using any of the actual components and assemblies produced in volume for current commercial players, it is likely the player would cost 1000–10 000 times as much as the current consumer price. A valid question is whether the ideas being researched could result in low enough cost if they were to be commercialized at reasonably large volume. It is debatable whether any approach that is based on evolutionary improvement of technologies developed for long-distance optical telecommunications can ever offer low enough cost; a pessimist might argue that is like asking vacuum tubes to evolve to make microprocessors. There are, however, revolutionary technologies that might offer sufficient economies of scale and could be manufacturable at low cost. An optimist would point to large two-dimensional arrays of optoelectronic devices hybridized to silicon CMOS, to the possibility of sophisticated molded plastic free-space optical assemblies handling thousands of light beams, and to the fact that optical fiber is already comparable in cost to electrical cables. A fair assessment might be that any estimate of final cost is currently highly speculative, and achieving low cost will require both considerable ingenuity and the promise of volume markets.

In conclusion, optics is a very promising technology for dense interconnects to silicon chips. It needs substantial technological work, and there are barriers, psychological, technological, and financial, to its introduction. Optics offers so many advantages, both qualitative and quantitative, that, if it were introduced, it could substantially alter information processing systems, and perhaps could become as indispensable on and between chips as it is today in long-distance communications.

Appendix
Bandwidth and Delay of Global On-Chip Electrical Interconnects

Here we summarize the models used to calculate the performance of on-chip lines as shown in Figs. 2 and 3. For these calculations, we start from standard approaches and introduce necessary extensions to those models. The extensions include 1) the effects of line inductance in limiting the performance of RC lines on chips and 2) consequences of the skin effect in limiting the prospects for LC lines on chips.

We consider explicitly the existing 0.25-μm and the projected 0.1-μm technology generations, which show the clear trends in the behavior. These trends continue in future projected generations. According to the SIA Roadmap [1], on-chip clock rates are $\sim$750 MHz for the 0.25-μm linewidth technology generation, and will rise to 3.5 GHz for the 0.1-μm generation. Chip size will rise only modestly from about 1.7 cm across to about 2.2 cm (hence we can presume $\sim 2 \times 2$ cm^2 chips for simplicity).

The global interconnect metal lines on the 0.25-μm generation chips are ~ 1 μm wide, and for simplicity we take them to be 1×1 μm in cross-section (actual lines are somewhat taller than this, e.g., 1.8 μm high, though this makes little difference to this calculation). We presume such a line takes up $\sim$8 square micrometers of cross-sectional area altogether because of the spacing between adjacent lines and metal levels, and the RC line bandwidth shown in Fig. 2 is the bandwidth deduced from (1). The bandwidth for an LC line (with the same dimensions) is similarly deduced from (2). The skin depth (the effective depth in which conduction can occur) is $\delta \cong 1/\sqrt{\pi f \mu_r \mu_o / \rho}$, where f is the frequency, μ_r is the relative permeability, μ_o is the permeability of free space, and ρ is the resistivity of the conducting material. For copper at 1 GHz, $\delta \cong 2.1$ μm. The crossover from RC-like to LC-like behavior (and also the crossover from bulk resistance to skin-effect resistance behavior) occurs near a frequency $\sim f_{\rm co}$ given by [4]

$$f_{\rm co} \cong \frac{17.5}{d^2_{(\mu{\rm m})}} \text{ GHz} \tag{4}$$

for copper lines, where d is the cross-sectional dimension of the (smaller or inner) conductor in the line (in micrometers). (Simulation results can suggest a somewhat lower frequency for the transition from RC to LC behavior, and a somewhat higher frequency for the onset of skin-effect resistance behavior.)

We start by following the modeling approach of Bakoglu [98] for repeatered lines. Such modeling presumes inductive effects are negligible, and assumes bulk line resistance (i.e., no skin effect); it also simply linearly adds the delays or rise times from different parts of a repeatered line segment—the internal RC time of the driver transistor, the effective RC time of the distributed RC line segment between the driver and receiver, and the RC time from charging the receiver transistor input capacitance through the distributed line resistance. Such a linear addition can underestimate actual delay and rise times by as much as a factor of two compared to exact simulations for the RC case, though adding inductive effects can somewhat speed up the rise time of the signal compared to this analysis.

The repeatered line is presumed to be designed for maximum velocity of propagation. The optimum properties are deduced by minimizing the total delay with respect to the driver size and the length of the repeatered line segment. For such lines, the propagation delay (50% rise time) on each segment of the line is set approximately equal to the delay (50% rise time) in each amplifier stage [98]. The repeater is made from transistors with a characteristic internal RC time constant $R_o C_o$; this time constant is approximately independent of the width of the transistor since resistance and capacitance scale oppositely as the transistor is made wider, and it is a characteristic of a given technology generation. The 50% rise time of the repeater alone is therefore $\cong 0.7 R_o C_o$. For a line segment with total (distributed) resistance and capacitance $R_{\rm dist}$ and $C_{\rm dist}$, respectively, the 50% rise time is $\cong 0.4 R_{\rm dist} C_{\rm dist}$ [98], so for the maximum velocity line,

$R_{\rm dist}C_{\rm dist} \cong 1.75R_oC_o$. (Note that the rise time of a distributed RC line is shorter than that of a lumped RC circuit because the "near" capacitance in the line is charged through a resistance $\ll R_{\rm dist}$.) There are additional delay terms for the charging of the transistor input capacitance through the line resistance ($0.7R_{\rm dist}C_o$, on the simple assumption that the receiver transistor input capacitance is equal to the driver transistor output capacitance, though in fact it will likely be less because the receiver transistor will likely be smaller), and the charging of the line capacitance through the transistor resistance ($0.7R_oC_{\rm dist}$). The optimum choice of the size of the transistor is one in which the actual output resistance is $R_o \cong \sqrt{R_oC_oR_{\rm dist}/C_{\rm dist}}$, and the actual output capacitance is $C_o = \sqrt{R_oC_oC_{\rm dist}/R_{\rm dist}}$, so each of these additional delay terms is $\cong 0.7\sqrt{R_oC_oR_{\rm dist}C_{\rm dist}} = 0.93R_oC_o$.

Hence, the total delay τ_r for a repeatered line segment is $\tau_r \cong 3.3R_oC_o = 1.9R_{\rm dist}C_{\rm dist}$. The length ℓ of a repeatered line segment is one such that $R_{\rm int}C_{\rm int}\ell^2 \equiv R_{\rm dist}C_{\rm dist} = 1.75R_oC_o$, where $R_{\rm int}$ and $C_{\rm int}$ are the resistance and capacitance per unit length of the interconnect, respectively. Hence $\ell \cong 1.32\sqrt{R_oC_o/R_{\rm int}C_{\rm int}}$. The resulting effective signal velocity is $v_r \cong \ell/\tau_r$, i.e.,

$$v_r \cong \frac{1}{2.5\sqrt{R_oC_oR_{\rm int}C_{\rm int}}}. \quad (5)$$

This effective velocity is, of course, an average quantity; the actual progress of the signal is not a smooth propagation down a line, but is more like a discrete set of steps in and out of repeater amplifiers and line segments. This effective velocity is useful for broad scaling arguments, however, and we use it to draw the repeatered line delay in Fig. 3.

Such a line is ready for another signal pulse when the whole repeater stage has settled. The repeatered line therefore behaves more like an LC line in that multiple pulses can be propagating down the entire line at once. (Note, though, that the entire length of the repeatered line is actively charged and discharged by the amplifier circuits for every pulse propagated, whereas in a true LC line a pulse of energy propagates down the line, so the power dissipation in a repeatered line is in general much higher.) The driver transistor 90% rise time is $\cong 2.3R_oC_o$, and the 90% rise time of a distributed line is $\cong R_{\rm dist}C_{\rm dist}$. As above, there are two other RC delay terms, which each are $\cong 2.3\sqrt{R_oC_oR_{\rm dist}C_{\rm dist}} = 3.1R_oC_o$, so the total 90% rise time on this simple model is $(2.3+3.1+1.75+3.1)R_oC_o \cong 10R_oC_o$. Hence the bandwidth of such a line is

$$B_r \cong \frac{1}{10R_oC_o}. \quad (6)$$

Note, incidentally, that this bandwidth is independent of the cross-sectional size of the line. A thicker line will have longer segments and a faster effective signal velocity, and a thinner line will have shorter segments and a slower effective signal velocity, but both will have the same bandwidth because this is set by the transistor parameters. This means, incidentally, that it is possible to get larger overall bandwidths by using more, thinner lines, though the propagation velocity suffers.

For our illustrative calculations here, we use an R_oC_o that is twice the "gate delay metric" quoted by SIA [1] for the different technology generations. This is a compromise between an optimistic view that might neglect all additional capacitances in a real gate and assume that the peak current flows at all voltages, and a more pessimistic heuristic that gives about four times the gate delay metric [3]. Here, therefore, we have $R_oC_o \cong 33$ ps for the 0.25-μm technology generation, and we would have$R_oC_o \cong 14$ ps for the 0.1-μm technology generation. This gives a bandwidth $B_r \cong 3$ GHz for the 0.25-μm technology generation, as plotted in Fig. 2.

For the 1-μm copper conductor, taking the bulk copper resistivity (an optimistic assumption for a thin line), we have $R_{\rm int} \cong 170\ \Omega$/cm, and we take $C_{\rm int} \cong 2$ pf/cm (somewhat lower values might be possible for future low dielectric constant materials). Using $R_oC_o \cong 33$ ps for the 0.25-μm technology, the velocity is $v_r \cong 3.78 \times 10^9$ cm/s $\cong 12.6\%\ c$, where c is the velocity of light in free space. This velocity is plotted as the delay of the 1-μm repeatered line in Fig. 3. (Slightly faster velocities may be possible in an optimized line design.) The delay in an individual repeatered line segment is 109 ps, which gives $\ell \cong 4.1$ mm for the 1×1 μm conductor cross-section we are considering.

An interesting question is how the propagation velocity of a repeatered line will change with technology generation. Suppose, first of all, that the above arguments remain valid as we go to future technology generations. Then there are two simple extremes.

i) We could suppose that we retain the same cross-section of upper interconnect line, which means we are likely substantially increasing the number of metal levels if we retain scaling ratios between metal line levels. In this case, since, semiempirically, $R_oC_o \propto s$ [1], where s is the technology generation size parameter (e.g., 0.25 or 0.1 μm), we should expect from (5) that $v_r \propto 1/\sqrt{s}$. This would give, for the 0.1-μm generation, $v_r \cong 20\%\ c$.

ii) We could alternatively suppose that the size of the upper interconnect metal scales down with the technology generation, so $R_{\rm int} \propto 1/s^2$ because of the decrease in line cross-sectional area. Taken together with $R_oC_o \propto s$, we arrive at $v_r \propto \sqrt{s}$, which represents an decrease of velocity with advancing technology generations.

Hence a compromise scaling would be halfway between these extremes, which would give a constant velocity in repeatered lines independent of technology generation, in which case the "1-μm repeatered line" delay would remain the delay in future technology generations.

A more subtle issue is the fact that the physical propagation velocity on the actual line segments themselves must be limited by the velocity of light; to understand these effects, we must move beyond the simple model of Bakoglu [98]. The propagation delay on the line itself is, on a simple distributed RC model, $\tau_\ell \cong 0.4R_{\rm dist}C_{\rm dist} = 0.4R_{\rm int}C_{\rm int}\ell^2$, corresponding to an effective velocity on the line itself of $v_\ell \cong \ell/\tau_\ell = 2.5/(R_{\rm int}C_{\rm int}\ell)$. This velocity will equal

$c/\sqrt{\varepsilon_r}$ for a length $\ell_c \cong 2.5\sqrt{\varepsilon_r}/(R_{\text{int}}C_{\text{int}}c)$ Presuming we make $\varepsilon_r \cong 2$, this point is reached for a line length of our 1-μm line of ~3.5 mm. In other words, we are very close already with the 0.25-μm technology to the limit of a simple RC repeatered line analysis and performance.

To model the lines properly so that such "velocity of light" effects are automatically included requires adding in the effects of line inductance. The inclusion of inductance and consequent LC transmission line effects is discussed in [91]–[97], where simulations show more detailed behavior of the cross-over from RC to RLC (a mixture of RC and LC) and LC behavior. Here we take a relatively simplistic analytic approach to expose the trends and underlying physical regimes. If we neglect the skin effect for the moment (which only makes line resistance effects even worse), the line can be modeled by the "telegraphers equation" for the voltage V

$$L_{\text{dist}}C_{\text{dist}}\frac{\partial^2 V}{\partial t^2} + R_{\text{dist}}C_{\text{dist}}\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2} \qquad (7)$$

where L_{dist} is the inductance per unit length (and where we presume no parallel conductance across the line). This equation can be changed into a dimensionless, universal form by choosing the unit of time as $t_o = L_{\text{dist}}/R_{\text{dist}}$ and the unit of length as $x_o = Z/R_{\text{dist}}$, where $Z = \sqrt{L_{\text{dist}}/C_{\text{dist}}}$ is the impedance of the "lossless" line.

If we drive such a line with a step function, a relatively sharp step does propagate down the line at approximately the velocity of light in the medium, $v_p = 1/\sqrt{L_{\text{dist}}C_{\text{dist}}}$. The amplitude of the propagating rising step does, however, attenuate, and a long rising tail follows behind the step. Numerical simulations show that the delay to 50% rise in the line is $\cong v_p\ell$ up to distances $\cong 2x_o$, (characteristic of propagation at v_p), but for longer distances, the propagating rising step is <50% of the drive voltage, and the delay changes to $\cong 0.4R_{\text{dist}}C_{\text{dist}}\ell^2$.

The inductance of most lines is L_{dist}~2–5 nH/cm (it is difficult to move far below this range because of the logarithmic dependence of inductance on the separation and size of conductors in lines). We use 2 nH/cm here for calculations. For the 1-μm line, we have been considering, therefore, we calculate $x_o \cong 1.87$ mm, so the lower limit of the simple RC charging model of line delay for this line is ~3.7 mm, similar to our heuristic estimate above. For line segments shorter than this distance, the line delay is *longer* than the RC charging model predicts, and so, for the 1-μm line, we cannot continue scaling with the simple RC charging model as technology generations advance.

To push to faster repeatered lines, we can see that we are also transitioning into the regime of transmission lines, where reflections and impedance matching become important. This could mean substantial changes in the drivers and receivers on the line and necessitate more sophisticated signaling (and likely larger latency in the line drivers and/or receivers).

These arguments taken together suggest that it will be difficult with repeatered lines to have delays much less than the 1-μm line/ 0.25-μm technology calculation shown in Fig. 3.

One could argue, of course, that if the delay of global interconnects is such a problem, then one will simply build thicker upper metal interconnect lines and connect without repeaters across the chip. This, however, requires quite large lines. For example, to get the rising step to propagate at >50% height across the chip requires $2x_o \cong 2$ cm, which requires $R_{\text{dist}} \sim 30$ Ω/cm; this, in turn, requires a cross-sectional area of copper conductor of ~6 μm^2 (assuming essentially all the resistance comes from only one conductor, i.e., no ground resistance), which corresponds to a ~2.4 × 2.4 μm^2 conductor. To make a transmission line will require a total cross-sectional area at least ~ten times as large (to allow sufficient space between neighboring conductors), for a total cross-sectional area of ~60 μm^2. These dimensions are significantly larger than current interconnect lines on silicon chips.

Furthermore, this global LC line will start to behave as a skin-effect limited line once the skin depth becomes significantly smaller than the thickness of the conductor. We might just expect to see such effects when the skin depth is about half the conductor thickness since there is penetration from both sides. At a frequency of ~3 GHz, which is about the clock frequency for the 0.1-μm generation, the skin depth is ~1.2 μm. Hence, the above hypothetical line with a 2.4 × 2.4 μm^2 (center) conductor should be able to carry a signal at the clock frequency across the chip (without equalization). If we consider the technology generations beyond 0.1 μm, with higher clock frequencies, the problem becomes worse. Low frequencies that are not limited by skin effect have less attentuation than high frequencies, and long tails can develop on the step response of such skin-effect limited lines. When the skin depth is much less than the line dimension, this effect can reduce the bandwidth of an unequalized line by ~×10 compared to the bandwidth of an RC line of the same dimensions [4]. For example, at 10-GHz clock, as envisaged for the 0.05-μm generation, skin depth is ~0.7 μm. The 2.4 × 2.4 μm^2 conductor in the hypothetical line would then be strongly into the skin effect limited regime and, according to (2), could have to have a total cross-sectional area as large as ~4000 μm^2 (≡63 × 63 μm^2) to carry a 10-GHz bandwidth across the chip. The detailed consequences of the skin effect should be analyzed by more detailed simulations, but this calculation illustrates that the skin effect could substantially increase the required size of lines for unequalized propagation across chips at future high clock speeds.

Hence we see that the consequences of line inductance and the skin effect provide substantial additional limits to the continued scaling of on-chip global electrical interconnect lines in future generations of silicon chips.

ACKNOWLEDGMENT

The author is pleased to acknowledge stimulating and informative conversations with R. Baets, B. Dally, T. Drabik,

G. Fitzgerald, M. Horowitz, K. Kimerling, J. Meindl, K. Saraswat, and S. Wong. He is also indebted to the reviewers of this paper for many constructive comments and suggestions.

REFERENCES

[1] "The National Technology Roadmap for Semiconductors—Technology Needs," Semiconductor Industry Association, 1997.

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

[3] M. Horowitz, C. K. K. Yang, and S. Sidiropoulos, "High-speed electrical signaling: Overview and limitations," *IEEE Micro*, vol. 18, no. 1, pp. 12–24, Jan./Feb. 1998.

[4] 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, pp. 42–52, 1997. (Special Issue on Parallel Computing with Optical Interconnects).

[5] J. W. Goodman, F. J. Leonberger, S.-Y. Kung, and R. A. Athale, "Optical interconnections for VLSI systems," *Proc. IEEE*, vol. 72, pp. 850–866, 1984.

[6] J. W. Goodman, "Fan-in and fan-out with optical interconnections," *Opt. Acta*, vol. 32, pp. 1489–1496, 1985.

[7] L. A. Bergman, W. H. Wu, A. R. Johnston, R. Nixon, S. C. Esener, C. C. Guest, P. Yu, T. J. Drabik, M. R. Feldman, and S. H. Lee, "Holographic optical interconnects for VLSI," *Opt. Eng.*, vol. 25, pp. 1109–1118, 1986.

[8] W. H. Wu, L. A. Bergman, A. R. Johnston, C. C. Guest, S. C. Esener, P. K. L. Yu, M. R. Feldman, and S. H. Lee, "Implementation of optical interconnections for VLSI," *IEEE Trans. Electron Devices*, vol. 34, pp. 706–714, 1987.

[9] M. R. Feldman, S. C. Esener, C. C. Guest, and S. H. Lee, "Comparison between optical and electrical interconnects based on power and speed considerations," *Appl. Opt.*, vol. 27, pp. 1742–1751, 1988.

[10] M. R. Feldman, C. C. Guest, T. J. Drabik, and S. C. Esener, "Comparison between electrical and free space optical interconnects for fine-grain processor arrays based on interconnect density capabilities," *Appl. Opt.*, vol. 28, pp. 3820–3829, 1989.

[11] 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.

[12] H. M. Ozaktas and J. W. Goodman, "Lower bound for the communication volume required for an optically interconnected array of points," *J. Opt. Soc. Amer.*, vol. A7, pp. 2100–2106, 1990.

[13] F. E. Kiamilev, P. Marchand, A. V. Krishnamoorthy, S. C. Esener, and S. H. Lee, "Performance comparison between optoelectronic and VLSI multistage interconnection networks," *J. Lightwave Technol.*, vol. 9, pp. 1674–1692, 1991.

[14] H. M. Ozaktas and J. W. Goodman, "Implications of interconnection theory for optical digital computing," *Appl. Opt.*, vol. 31, pp. 5559–5567, 1992.

[15] ——, "Elements of a hybrid interconnection theory," *Appl. Opt.*, vol. 33, pp. 2968–2987, 1994.

[16] T. J. Drabik, "Optoelectronic integrated systems based on free-space interconnects with an arbitrary degree of space variance," *Proc. IEEE*, vol. 82, pp. 1595–1622, 1994.

[17] L. J. Camp, R. Sharma, and M. R. Feldman, "Guided-wave and free-space optical interconnects for parallel-processing systems: A comparison," *Appl. Opt.*, vol. 33, pp. 6168–6180, 1994.

[18] C. Fan, B. Mansoorian, D. A. Vanblerkom, M. W. Hansen, V. H. Ozguz, S. C. Esener, and G. C. Marsden, "Digital free-space optical interconnections: A comparison of transmitter technologies," *Appl. Opt.*, vol. 34, no. 17, pp. 3103–3115, June 10, 1995.

[19] T. Nakahara, S. Matsuo, S. Fukushima, and T. Kurokawa, "Performance comparison between multiple-quantum-well modulator-based and vertical-cavity-surface-emitting laser-based smart pixels," *Appl. Opt.*, vol. 35, pp. 860–871, 1996.

[20] F. A. P. Tooley, "Challenges in optically interconnecting electronics," *IEEE J. Select. Topics Quantum Electron.*, vol. 2, pp. 3–13, 1996.

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

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

[23] H. M. Ozaktas, "Toward an optimal foundation architecture for optoelectronic computing. 1. Regularly interconnected devices planes," *Appl. Opt.*, vol. 36, pp. 5682–5696, 1997.

[24] ——, "Toward an optimal foundation architecture for optoelectronic computing: 2. Physical construction and application platforms," *Appl. Opt.*, vol. 36, pp. 5697–5705, 1997.

[25] D. A. Van Blerkom, C. Fan, M. Blume, and S. C. Esener, "Transimpedance receiver design optimization for smart pixel arrays," *J. Lightwave Technol.*, vol. 16, pp. 119–126, 1998.

[26] G. I. Yayla, P. J. Marchand, and S. C. Esener, "Speed and energy analysis of digital interconnections: Comparison of on-chip, off-chip, and free-space technologies," *Appl. Opt.*, vol. 37, pp. 205–227, 1998.

[27] A. V. Krishnamoorthy, "Applications of optoelectronic-VLSI technologies," *Int. J. Optoelectron.*, vol. 12, pp. 155–161, 1998.

[28] J. Jahns, "Free-space optical digital computing and interconnection," *Progress Opt.*, vol. 38, pp. 419–513, 1998.

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

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

[31] J. Tanida, K. Nitta, T. Inoue, and Y. Ichioka, "Comparison of electrical and optical interconnection for large fan-out communication," *J. Opt. A*, vol. 1, pp. 262–266, 1999.

[32] E. E. E. Frietman, F. Zhao, and G. D. Khoe, "A prototype for optical interconnection in massively parallel processing and its physical and optical modeling," *J. Opt. A*, vol. 1, pp. 290–294, 1999.

[33] T. J. Drabik, "Balancing electrical and optical interconnection resources at low levels," *J. Opt. A*, vol. 1, pp. 330–332, 1999.

[34] A. L. Lentine and D. A. B. Miller, "Evolution of the SEED technology: Bistable logic gates to optoelectronic smart pixels," *IEEE J. Quantum Electron.*, vol. 29, pp. 655–669, 1993.

[35] H. S. Hinton, "Progress in the smart pixel technologies," *IEEE J. Select. Topics Quantum Electron.*, vol. 2, pp. 14–23, 1996.

[36] 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.

[37] D. A. B. Miller, D. S. Chemla, T. C. Damen, T. H. Wood, C. A. Burrus, A. C. Gossard, and W. Wiegmann, "The quantum well self-electrooptic effect device: Optoelectronic bistability and oscillation, and self linearized modulation," *IEEE J. Quantum Electron.*, vol. QE-21, pp. 1462–1476, 1985.

[38] D. A. B. Miller, "Quantum-well self-electro-optic effect devices," *Opt. Quantum Electron.*, vol. 22, pp. S61–S98, 1990.

[39] E. R. Hegblom, N. M. Margalit, A. Fiore, and L. A. Coldren, "Small efficient vertical-cavity lasers with tapered oxide apertures," *Electron. Lett.*, vol. 34, pp. 895–897, 1998.

[40] A. V. Krishnamoorthy, L. M. F. Chirovsky, W. S. Hobson, R. E. Leibenguth, S. P. Hui, C. J. Zydzik, K. W. Goossen, J. D. Wynn, B. J. Tseng, J. Lopata, J. A. Walker, J. E. Cunningham, and L. A. D'Asaro, "Vertical-cavity surface-emitting lasers flip-chip bonded to gigabit-per-second CMOS circuits," *IEEE Photon. Technol. Lett.*, vol. 11, pp. 128–130, 1999.

[41] 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.

[42] 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 symmetrical self-electro-optic-effect devices," *Appl. Opt.*, vol. 32, pp. 5153–5171, 1993.

[43] M. J. Goodwin, A. J. Moseley, M. Q. Kearley, R. C. Morris, C. J. G. Kirkby, J. Thompson, R. C. Goodfellow, and I. Bennion, "Optoelectronic component arrays for optical interconnection of circuits and subsystems," *J. Lightwave Technol.*, vol. 9, pp. 1639–1645, 1991.

[44] 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.

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

[46] D. A. B. Miller, "Dense two-dimensional integration of optoelectronics and electronics for interconnections," in *Heterogeneous Integration: Systems on a Chip*, A. Husain and M. Fallahi, Eds. Bellingham, WA, 1998, vol. CR70, SPIE Critical Reviews of Optical Engineering, pp. 80–109.

[47] A. Krishnamoorthy, A. L. Lentine, K. W. Goossen, J. A. Walker, T. K. Woodward, J. E. Ford, G. F. Aplin, L. A. D'Asaro, S. P. Hui, B. Tseng, R. Leibenguth, D. D. Kossives, D. Dahringer, L. M. F. Chirovsky, and D. A. B. Miller, "3-D integration of MQW modulators over active submicron CMOS circuits: 375 Mb/s transimpedance receiver-transmitter circuit," *IEEE Photon. Technol. Lett.*, vol. 7, pp. 1288–1290, 1995.

[48] A. L. Lentine, K. W. Goossen, J. A. Walker, J. E. Cunningham, W. Y. Jan, T. K. Woodward, A. V. Krishnamoorthy, B. J. Tseng, S. P. Hui, R. E. Leibenguth, L. M. F. Chirovsky, R. A. Novotny, D. B. Buchholz, and R. L. Morrison, "Optoelectronic VLSI switching chip with >1 Tbit/s potential optical I/O bandwidth," *Electron. Lett.*, vol. 33, pp. 894–895, 1997.

[49] M. Kufner and S. Kufner, *Micro-optics and Lithography*. Brussels, Belgium: VUB Press, 1997.

[50] H. S. Hinton, T. J. Cloonan, F. B. McCormick, A. L. Lentine, and F. A. P. Tooley, "Free-space digital optical-systems," *Proc. IEEE*, vol. 82, pp. 1632–1649, 1994.

[51] S. Sinzinger and J. Jahns, "Integrated micro-optical imaging-system with a high interconnection capacity fabricated in planar optics," *Appl. Opt.*, vol. 36, pp. 4729–4735, 1997.

[52] D. T. Neilson and E. Schenfeld, "Free-space optical relay for the interconnection of multimode fibers," *Appl. Opt.*, vol. 38, pp. 2291–2296, 1999.

[53] G. Verschaffelt, R. Buczynski, P. Tuteleers, P. Vynck, V. Baukens, H. Ottevaere, C. Debaes, S. Kufner, M. Kufner, A. Hermanne, J. Genoe, D. Coppee, R. Vounckx, S. Borghs, I. Veretennicoff, and H. Thienpont, "Demonstration of a monolithic multichannel module for multi-Gb/s intra-MCM optical interconnects," *IEEE Photon. Technol. Lett.*, vol. 10, pp. 1629–1631, 1998.

[54] H. Kosaka, M. Kajita, Y. Li, and Y. Sugimoto, "A 2-dimensional optical parallel transmission using a vertical-cavity surface-emitting laser array module and an image fiber," *IEEE Photon. Technol. Lett.*, vol. 9, pp. 253–255, 1997.

[55] B. Robertson, D. Kabal, G. C. Boisset, Y. S. Liu, W. M. Robertson, M. R. Taghizadeh, and D. V. Plant, "Design and operation of an in-situ microchannel alignment-detection system," *Appl. Opt.*, vol. 37, pp. 5368–5376, 1998.

[56] D. V. Plant, "Packaging techniques for smart pixel optoelectronics," *Int. J. Optoelectron.*, vol. 11, pp. 169–180, 1997.

[57] M. C. Wu, "Micromachining for optical and optoelectronic systems," *Proc. IEEE*, vol. 85, pp. 1833–1856, 1997.

[58] A. C. Walker, T. Y. Yang, J. Gourlay, J. A. B. Dines, M. G. Forbes, S. M. Prince, D. A. Baillie, D. T. Neilson, R. Williams, L. C. Wilkinson, G. R. Smith, M. P. Y. Desmulliez, G. S. Buller, M. R. Taghizadeh, A. Waddie, I. Underwood, C. R. Stanley, F. Pottier, B. Vogele, and W. Sibbett, "Optoelectronic systems based on InGaAs-complementary-metal-oxide-semiconductor smart-pixel arrays and free-space optical interconnects," *Appl. Opt.*, vol. 37, pp. 2822–2830, 1998.

[59] M. H. Ayliffe, D. Kabal, F. Lacroix, E. Bernier, P. Khurana, A. G. Kirk, F. A. P. Tooley, and D. V. Plant, "Electrical, thermal and optomechanical packaging of large 2D optoelectronic device arrays for free-space optical interconnects," *J. Opt. A*, vol. 1, pp. 267–271, 1999.

[60] A. M. Agarwal, L. Liao, J. S. Foresi, M. R. Black, X. M. Duan, and L. C. Kimerling, "Low-loss polycrystalline silicon wave-guides for silicon photonics," *J. Appl. Phys.*, vol. 80, pp. 6120–6123, 1996.

[61] B. Jalali, P. D. Trinh, S. Yegnanarayanan, and F. Coppinger, "Guided-wave optics in Silicon-on-insulator technology," *Proc. Inst. Elect. Eng. Optoelectron.*, vol. 143, pp. 307–311, 1996.

[62] J. S. Foresi, D. R. Lim, A. M. Agarwal, L. C. Kimerling, M. Tavassoli, M. Cox, M. Cao, and W. Greene, "Small radius bends and large angle splitters in SOI waveguides," in *Proc. SPIE "Photonics West" Symp.*, vol. 3007, San Jose, CA, Feb. 1997, pp. 112–118.

[63] B. D. Clymer and J. W. Goodman, "Optical clock distribution to silicon chips," *Opt. Eng.*, vol. 25, pp. 1103–1108, 1986.

[64] ——, "Timing uncertainty for receivers in optical clock distribution for VLSI," *Opt. Eng.*, vol. 27, pp. 944–954, 1988.

[65] P. J. Delfyett, D. H. Hartman, and S. Z. Ahmad, "Optical clock distribution using a mode-locked semiconductor-laser diode system," *J. Lightwave Technol.*, vol. 9, pp. 1646–1649, 1991.

[66] S. J. Walker and J. Jahns, "Optical clock distribution using integrated free-space optics," *Opt. Commun.*, vol. 90, pp. 359–371, 1992.

[67] G. Farrell, P. Phelan, and J. Hegarty, "All-optical clock distribution with synchronous frequency-division and multiplication," *Electron. Lett.*, vol. 28, pp. 1387–1388, 1992.

[68] A. Aboudou, J. P. Vilcot, G. Danneville, D. Decoster, E. Delhaye, P. Boissenot, C. Varin, F. Deschamps, and I. Lecur, "GaAlAs/GaAs planar photoconductors and MSM photodetectors monolithically integrated with HIGFETS: Application for optical clock distribution," *Proc. Inst. Elect. Eng. Optoelectron.*, pt. J, vol. 139, pp. 83–87, 1992.

[69] T. Schwider, N. Streibl, T. Schmidt, K. Zurl, and R. Volkel, "Synchronous optical clock distribution for optoelectronic interconnections," *Opt. Lett.*, vol. 19, pp. 75–77, 1994.

[70] S. N. Tang and R. T. Chen, "1-to-42 optoelectronic interconnection for intra-multichip-module clock signal distribution," *Appl. Phys. Lett.*, vol. 64, pp. 2931–2933, 1994.

[71] H. Zarschizky, C. Gerndt, M. Honsberg, H. W. Schneider, and A. Stemmer, "Optical clock distribution with a compact free-space interconnect system," *Opt. Quantum Electron.*, vol. 26, pp. S471–S481, 1994.

[72] S. Koh, H. W. Carter, and J. T. Boyd, "Synchronous global clock distribution on multichip modules using optical wave-guides," *Opt. Eng.*, vol. 33, pp. 1587–1595, 1994.

[73] L. S. Tamil and K. Hsia, "An energy-efficient free-space optical interconnection for clock distribution," *Microwave Opt. Technol. Lett.*, vol. 9, pp. 19–22, 1995.

[74] J. H. Yeh, R. K. Kostuk, and K. Y. Tu, "Board-level H-tree optical clock distribution with substrate mode holograms," *J. Lightwave Technol.*, vol. 13, pp. 1566–1578, 1995.

[75] K. S. Lee and C. Shu, "Optical matrix for clock distribution and synchronous operation in 2-dimensional array devices," *Appl. Phys. Lett.*, vol. 68, pp. 3528–3530, 1996.

[76] R. T. Chen, F. M. Li, Y. M. Dubinovsky, and O. Ershov, "Si-based surface-relief polygonal gratings for 1-to-many wafer-scale optical clock signal distribution," *IEEE Photon. Technol. Lett.*, vol. 8, pp. 1038–1040, 1996.

[77] M. P. Y. Desmulliez, P. W. Foulk, and M. R. Taghizadeh, "Optical clock distribution for multichip-module," *Opt. Rev.*, vol. 3, pp. 379–380, 1996.

[78] B. Lunitz and J. Jahns, "Tolerant design of a planar-optical clock distribution-system," *Opt. Commun.*, vol. 134, pp. 281–288, 1997.

[79] S. K. Tewksbury and L. A. Hornak, "Optical clock distribution in electronic systems," *J. VLSI Signal Process. Syst. Signal Image Video Technol.*, vol. 16, pp. 225–246, 1997.

[80] C. H. Zhao and R. T. Chen, "Performance consideration of 3-dimensional optoelectronic interconnection for intra-multichip-module clock signal distribution," *Appl. Opt.*, vol. 36, pp. 2537–2544, 1997.

[81] Y. Li, T. Wang, J. K. Rhee, and L. J. Wang, "Multigigabits per second board-level clock distribution schemes using laminated end-tapered fiber-bundles," *IEEE Photon. Technol. Lett.*, vol. 10, pp. 884–886, 1998.

[82] Y. Li, J. Popelek, L. J. Wang, Y. Takiguchi, T. Wang, and K. Shum, "Clock delivery using laminated polymer fiber circuits," *J. Opt. A*, vol. 1, pp. 239–243, 1999.

[83] D. Litaize, M. P. Y. Desmulliez, J. H. Collet, and P. Foulk, "Optical clock distribution for a more efficient use of DRAMs," *J. Opt. A*, vol. 1, pp. 295–298, 1999.

[84] B. J. Smith, "Interconnection networks for shared memory parallel computers," in *Proc. 2nd Int. Conf. Massively Parallel Processing Using Optical Interconnections (MPPOI'95)*. Los Alamitos, CA: IEEE Computer Society Press, pp. 255–256.

[85] G. A. Sai-Halasz, "Performance trends in high-end processors," *Proc. IEEE*, vol. 83, pp. 20–36, 1995.

[86] J. D. Meindl, "Low power microelectronics: Retrospect and prospect," *Proc. IEEE*, vol. 83, pp. 619–635, 1995.

[87] M. T. Bohr and Y. A. El-Mansy, "Technology for advanced high-performance microprocessors," *IEEE Trans. Electron Devices*, vol. 45, pp. 620–625, 1998.

[88] J. A. Davis and J. D. Meindl, "Is interconnect the weak link?," *IEEE Circuits Devices Mag.*, vol. 14, no. 2, pp. 30–36, Mar. 1998.

[89] J. A. Davis, V. K. De, and J. D. Meindl, "A stochastic wire-length distribution for gigascale integration GSI, I: Derivation and validation," *IEEE Trans. Electron Devices*, vol. 45, pp. 580–589, 1998.
[90] ——, "A stochastic wire-length distribution for gigascale integration GSI, II: Applications to clock frequency, power dissipation, and chip size estimation," *IEEE Trans. Electron Devices*, vol. 45, pp. 590–597, 1998.
[91] A. Deutsch, G. V. Kopcsay, P. J. Restle, H. H. Smith, G. Katopis, W. D. Becker, P. W. Coteus, C. W. Surovic, B. J. Rubin, R. P. Dunne Jr., T. Gallo, K. A. Jenkins, L. M. Terman, R. H. Dennard, G. A. Sai-Halasz, B. L. Krauter, and D. R. Knebel, "When are transmission-line effects important for on-chip interconnections?," *IEEE Trans. Microwave Theory Tech.*, vol. 45, pp. 1836–1846, 1997.
[92] A. Deutsch, G. V. Kopcsay, C. W. Surovic, B. J. Rubin, L. M. Terman, R. P. Dunne Jr., T. A. Gallo, and R. H. Dennard, "Modeling and characterization of long on-chip interconnections for high-performance microprocessors," *IBM J. Res. Dev.*, vol. 39, pp. 547–568, 1995.
[93] P. J. Restle, K. A. Jenkins, A. Deutsch, and P. W. Cook, "Measurement and modeling of on-chip transmission line effects in a 400 MHz microprocessor," *IEEE J. Solid-State Circuits*, vol. 33, pp. 662–665, 1998.
[94] A. Deutsch, H. Smith, G. V. Kopesay, D. C. Edelstein, and P. W. Coteus, "On-chip wiring design challenges for Ghz operation," in *Proc. Electrical Performance of Electronic Packaging*, Piscataway, NJ, Oct. 25–27, 1999, pp. 45–48.
[95] A. Deutsch, "Electrical characteristics of interconnections for high-performance systems," *Proc. IEEE*, vol. 86, pp. 315–357, 1998.
[96] A. Deutsch, H. Smith, G. A. Katopis, W. D. Becker, P. W. Coteus, C. W. Surovic, G. V. Kopcsay, B. J. Rubin, R. P. Dunne, T. Gallo, D. R. Knebel, B. L. Krauter, L. M. Terman, G. A. Sai-Halasz, and P. J. Reslte, "The importance of inductance and inductive coupling for on-chip wiring," in *Proc. IEEE 6th Topical Meeting Electrical Performance of Electronic Packaging*, Piscataway, NJ, Oct. 27–29, 1997, pp. 53–56.
[97] Y. I. Ismail, E. G. Friedman, and J. L. Neves, "Figures of merit to characterize the importance of on-chip inductance," *IEEE Trans. VLSI Syst.*, vol. 7, pp. 442–449, 1999.
[98] H. B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.
[99] S. Afonso, L. W. Schaper, J. P. Parkerson, W. D. Brown, S. S. Ang, and H. A. Naseem, "Modeling and electrical analysis of seamless high off-chip connectivity (SHOCC) interconnects," *IEEE Trans. Adv. Packag.*, vol. 22, pp. 309–320, 1999.
[100] 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, June 2000.
[101] M. G. Forbes and A. C. Walker, "Wide-band transconductance-transimpedance post-amplifier for large photoreceiver arrays," *Electron. Lett.*, vol. 34, pp. 589–590, 1998.
[102] M. Kuijk, D. Coppee, and R. Vounckx, "Spatially modulated light detector in CMOS with sense-amplifier receiver operating at 180 Mb/s for optical data link applications and parallel optical interconnects between chips," *IEEE J. Select. Topics Quantum Electron.*, vol. 4, pp. 1040–1045, 1998.
[103] T. K. Woodward, A. V. Krishnamoorthy, A. L. Lentine, and L. M. F. Chirovsky, "Optical receivers for optoelectronic VLSI," *IEEE J. Select. Topics Quantum Electron.*, vol. 2, pp. 106–116, 1996.
[104] ——, "1 Gbit/s CMOS photoreceiver with integrated detector operating at 850 nm," *Electron. Lett.*, vol. 34, pp. 1252–1253, 1998.
[105] A. V. Krishnamoorthy and D. A. B. Miller, "Firehose architectures for free-space optically interconnected VLSI circuits," *J. Parallel Distribut. Comput.*, vol. 41, pp. 109–114, 1997.
[106] Y. Yamamoto and H. A. Haus, "Preparation, measurement and information capacity of optical quantum states," *Rev. Mod. Phys.*, vol. 58, pp. 1001–1020, 1986.
[107] A. L. Lentine, K. W. Goossen, J. A. Walker, L. M. F. Chirovsky, L. A. D'Asaro, S. P. Hui, B. T. Tseng, R. E. Leibenguth, D. P. Kossives, D. W. Dahringer, D. D. Bacon, T. K. Woodward, and D. A. B. Miller, "Arrays of optoelectronic switching nodes comprised of flip-chip-bonded MQW modulators and detectors on silicon CMOS circuitry," *IEEE Photon. Technol. Lett.*, vol. 8, pp. 221–223, 1996.
[108] Y. K. Chen, M. C. Wu, T. Tanbunek, R. A. Logan, and M. A. Chin, "Subpicosecond monolithic colliding-pulse mode-locked multiple quantum-well lasers," *Appl. Phys. Lett.*, vol. 58, pp. 1253–1255, 1991.
[109] L. Boivin, M. C. Nuss, J. Shah, D. A. B. Miller, and H. A. Haus, "Receiver sensitivity improvement by impulsive coding," *IEEE Photon. Technol. Lett.*, vol. 9, no. 5, pp. 684–686, 1997.
[110] E. A. De Souza, M. C. Nuss, W. H. Knox, and D. A. B. Miller, "Wavelength-division multiplexing with femtosecond pulses," *Opt. Lett.*, vol. 20, pp. 1166–1168, 1995.
[111] G. F. Williams, "Lightwave receivers," in *Topics in Lightwave Transmission Systems*, T. Li, Ed. San Diego, CA: Academic, 1991.
[112] O. Sjolund, D. A. Louderback, E. R. Hegblom, J. Ko, and L. A. Coldren, "Monolithic integration of substrate input output resonant photodetectors and vertical-cavity lasers," *IEEE J. Quantum Electron.*, vol. 35, pp. 1015–1023, 1999.
[113] D. J. Goodwill, K. E. Devenport, and H. S. Hinton, "An ATM-based intelligent optical backplane using CMOS-SEED smart pixel arrays and free-space optical interconnect modules," *IEEE J. Select. Topics Quantum Electron.*, vol. 2, pp. 85–96, 1996.
[114] K. W. Goossen, J. E. Cunningham, and W. Y. Jan, "Stacked-diode electroabsorption modulator," *IEEE Photon. Technol. Lett.*, vol. 6, pp. 936–938, 1994.
[115] K. W. Goossen, J. E. Cunningham, W. Y. Jan, and D. A. B. Miller, "Interleaved-contact electroabsorption modulator using doping-selective electrodes with 25-degrees-C to 95-degrees-C operating range," *IEEE Photon. Technol. Lett.*, vol. 5, pp. 181–183, 1993.
[116] P. Zouganeli, P. J. Stevens, D. Atkinson, and G. Parry, "Design trade-offs and evaluation of the performance attainable by GaAs–$Al_{0.3}Ga_{0.7}As$ asymmetric Fabry–Perot modulator," *IEEE J. Quantum Electron.*, vol. 31, pp. 927–943, 1995.
[117] D. L. Huffaker, D. G. Deppe, K. Kumar, and T. J. Rogers, "Native-oxide defined ring contact for low-threshold vertical-cavity lasers," *Appl. Phys. Lett.*, vol. 65, pp. 97–99, 1994.
[118] D. M. Cutrer and K. Y. Lau, "Ultralow power optical interconnect with zero-biased, ultralow threshold laser—How low a threshold is low enough?," *IEEE Photon. Technol. Lett.*, vol. 7, pp. 4–6, 1995.
[119] J. Martinregalado, F. Prati, M. Sanmiguel, and N. B. Abraham, "Polarization properties of vertical-cavity surface-emitting lasers," *IEEE J. Quantum Electron.*, vol. 33, pp. 765–783, 1997.
[120] H. Uenohara, K. Tateno, T. Kagawa, Y. Ohiso, H. Tsuda, T. Kurokawa, and C. Amano, "Polarization-controlled 850-nm-wavelength vertical-cavity surface-emitting lasers grown on (311)B substrates by metal-organic chemical vapor deposition," *IEEE J. Select. Topics Quantum Electron.*, vol. 5, pp. 537–545, 1999.
[121] H. Deneve, J. Blondelle, P. Vandaele, P. Demeester, R. Baets, and G. Borghs, "Recycling of guided mode light emission in planar microcavity light-emitting diodes," *Appl. Phys. Lett.*, vol. 70, pp. 799–801, 1997.
[122] R. Bockstaele, J. Derluyn, C. Sys, S. Verstuyft, I. Moerman, P. VanDaele, and R. Baets, "Realization of highly efficient 850nm top emitting resonant cavity light emitting diodes," *Electron. Lett.*, vol. 35, pp. 1564–1565, 1999.
[123] K. D. Hirschman, L. Tsybeskov, S. P. Duttagupta, and P. M. Fauchet, "Silicon-based visible light-emitting devices integrated into microelectronic circuits," *Nature*, vol. 384, pp. 338–341, 1996.
[124] Z. H. Lu, D. J. Lockwood, and J. M. Baribeau, "Quantum confinement and light-emission in SiO_2/Si superlattices," *Nature*, vol. 378, pp. 258–260, 1995.
[125] D. Leong, M. Harry, K. J. Reeson, and K. P. Homewood, "A silicon/iron-disilicide light-emitting diode operating at a wavelength of 1.5 μm," *Nature*, vol. 387, pp. 686–688, 1997.
[126] T. Makimura, Y. Kunii, and K. Murakami, "Light-emission from nanometer-sized silicon particles fabricated by the laser-ablation method," *Jpn. J. Appl. Phys.*, vol. 35, pp. 4780–4784, 1996.
[127] L. Vervoort, F. Bassani, I. Mihalcescu, J. C. Vial, and F. A. Davitaya, "Efficient visible-light emission from Si/Caf2(111) heterostructures grown by molecular-beam epitaxy," *Phys. Status Solidi B*, vol. 190, pp. 123–127, 1995.
[128] A. Polman, "Erbium implanted thin-film photonic materials," *J. Appl. Phys.*, vol. 82, pp. 1–39, 1997.
[129] Y. S. Tang, S. E. Hicks, W. X. Ni, C. M. S. Torres, G. V. Hansson, and C. D. W. Wilkinson, "Controlling the strain and light-emission from Si–$Si_{1-x}Ge_x$ quantum dots," *Thin Solid Films*, vol. 294, pp. 304–307, 1997.
[130] J. Y. Zhang, Y. H. Ye, and X. L. Tan, "Electroluminescence and carrier transport of SiO2 film containing different density of Ge nanocrystals," *Appl. Phys. Lett.*, vol. 74, pp. 2459–2461, 1999.

[131] L. C. Kimerling, K. D. Kolenbrander, J. Michel, and J. Palm, "Light-emission from silicon," *Solid State Phys.*, vol. 50, pp. 333–381, 1997.

[132] G. V. Treyz, P. G. May, and J. M. Halbout, "Silicon Mach–Zehnder wave-guide interferometers based on the plasma dispersion effect," *Appl. Phys. Lett.*, vol. 59, pp. 771–773, 1991.

[133] U. Keller, "Semiconductor nonlinearities for solid-state laser modelocking and Q-switching," *Semiconduct. Semimetals*, vol. 59, pp. 211–286, 1999.

[134] S. Gee, G. Alphonse, J. Connolly, and P. J. Delfyett, "High-power mode-locked external-cavity semiconductor-laser using inverse bow-tie semiconductor optical amplifiers," *IEEE J. Select. Topics Quantum Electron.*, vol. 4, pp. 209–215, 1998.

[135] 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.

[136] R. M. Sieg, J. A. Carlin, J. J. Boeckl, S. A. Ringel, M. T. Currie, S. M. Ting, T. A. Langdo, G. Taraschi, E. A. Fitzgerald, and B. M. Keyes, "High minority-carrier lifetimes in GaAs grown on low-defect-density Ge/GeSi/Si substrates," *Appl. Phys. Lett.*, vol. 73, pp. 3111–3113, 1998.

[137] D. L. Mathine, "The integration of III–V optoelectronics with silicon circuitry," *IEEE J. Select. Topics. Quantum Electron.*, vol. 3, pp. 952–959, 1997.

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, in 1979.

He worked at Bell Laboratories from 1981 to 1996, being a Department Head from 1987, latterly of the Advanced Photonics Research Department. He is currently the W. M. Keck Foundation Professor of electrical engineering at Stanford University, Stanford, CA, and Director of the Ginzton and Solid State and Photonics Laboratories. His research interests include quantum-well optoelectronic physics and devices and fundamentals and applications of optics in information switching and processing. He has published more than 200 scientific papers and has received 40 patents.

Prof. Miller is a Fellow of the Royal Society of London, OSA, and APS. He received an honorary degree from Vrije Univeriteit Brussel, Belgium. He has served as a Board Member for both OSA and the IEEE Lasers and Electro-Optics Society and in various other society and conference committees. He was President of the IEEE Lasers and Electro-Optics Society in 1995. He received the Adolph Lomb Medal and R. W. Wood Prize from the OSA, the International Prize in Optics from the International Commission for Optics, and an IEEE Third Millennium Medal.