

Latency Reduction in Optical Interconnects Using Short Optical Pulses

Diwakar Agarwal, *Student Member, IEEE*, Gordon A. Keeler, *Student Member, IEEE*,
Christof Debaes, *Student Member, IEEE*, Bianca E. Nelson, *Student Member, IEEE*,
Noah C. Helman, *Student Member, IEEE*, and David A. B. Miller, *Fellow, IEEE*

Abstract—We present a new method of latency reduction in optical interconnects: using very low duty cycle return-to-zero encoding (i.e., subpicosecond pulses). An analytical comparison of three different receiver architectures, including transimpedance, integrating, and totem-pole diode pair, is presented. For all three receivers, we demonstrate that using short pulses instead of nonreturn-to-zero (NRZ) shortens the circuit delay. We also experimentally demonstrate a $\sim 65\%$ reduction in latency of a transimpedance receiver by using short optical pulses. Finally, we show that the latency of optical interconnects can be comparable to or even less than electrical interconnects for global on-chip communication.

Index Terms—Complementary metal–oxide–semiconductor (CMOS) integrated circuits, integrating receivers, latency, multiple quantum well (MQW) modulator, optical interconnects, short optical pulses, totem-pole diode pair, trans-impedance receivers.

I. INTRODUCTION

MODERN computer processors run at the clock speeds of many gigahertz but the processor to memory interface typically runs at only a few hundred megahertz. A key reason for this difference, and a problem for computing in general, is that the interface connection speeds are not able to keep up with the increase in the processor speeds. This is mainly because of design issues of electrical busses and their underlying physical properties. Due to the capacity limitations of electrical wires, most long distance communication is now done via optics. For medium distance communication, e.g., local-area network (LAN), metropolitan area networks (MAN), wide area network (WAN) (about 300 m–100 km), optics is making inroads, specifically, because only optics can support the high data rates required by these applications. At shorter distances (a few meters—few hundred meters), primarily in data links, optics is rapidly gaining entry. Recently, a lot of research has been done

on technologies for massively parallel optical interconnects, including dense integration of modulators [1] and vertical cavity surface emitting lasers (VCSELs) [2]. This might allow the use of optical interconnects in short distance inter-chip and even intra-chip communication.

In connections between and within electronic chips, signal latency is a critical parameter in determining system performance. As the complementary metal–oxide–semiconductor (CMOS) linewidth scales, the processor clock speed increases, making it difficult to run an entire chip synchronously. In other words, transferring data within a clock cycle is becoming difficult. According to the International Technology Roadmap for Semiconductors (ITRS) estimate [3], gate delay and local interconnect delay are being reduced as the technology is scaling, but the delay of global interconnects with and without repeaters is continuously increasing relative to the clock period.

The propagation velocity of global interconnects with repeaters is a small fraction of the velocity of light (e.g., $0.1c$ – $0.2c$) and is not expected to improve significantly [4]–[6]. For $0.25\ \mu\text{m}$ technology, the delay of global lines is less than a clock cycle, but for future technologies the delay will be longer than a clock cycle. If the signals can be propagated at a significant fraction of the velocity of light, (e.g., $>0.3c$), the delay in communication will be less than a clock cycle up to $0.1\ \mu\text{m}$ technology [4].

It might be possible to use optics to provide communication across chips at a significant fraction of the velocity of light. For optics to be feasible, the delay in the transmitter and the receiver has to be very low, of the order of a few gate delays. The delay of propagation in optical media cannot be altered though it is relatively fast ($\sim 0.67c$ in glass). Transmitter and receiver circuits designed in silicon CMOS are likely to keep pace with the speed of logic operations in silicon chips as the technology scales [7].

Latency is very important in optical interconnects as mentioned by Dambre *et al.* [8], who show that with low latency optical links, three-dimensional (3-D) optoelectronic multi-FPGAs outperform two-dimensional (2-D) electronic FPGAs. Also, recently Collet *et al.* [9] concluded that since the most critical issue in computer architecture is the access time to the main memory, signal latency is of critical importance in implementing optical interconnects. There is some concern, however, about the increased latency of optical interconnects compared with their electrical counterpart because of the added functions of electrical-to-optical and optical-to-electrical conversion as expressed in [10]. But because of advanced integration techniques [1], optical components with very low parasitics can be placed close to the origin of signal, which

Manuscript received November 1, 2002; revised February 7, 2003. This work was supported in part by the Defense Advanced Research Projects Agency under the University Optocenters Program under Contract MDA972-00-1-0024 through the University of New Mexico, and in part by DARPA/MARCO under the Interconnect Focus Center through the Georgia Institute of Technology. The work of C. Debaes was supported in part by a Fellowship from the FWO Vlaanderen. The work of N. Helman was supported in part by a Gerhard Casper Stanford Graduate Fellowship.

D. Agarwal, G. A. Keeler, B. E. Nelson, N. C. Helman, and D. A. B. Miller are with the Departments of Electrical Engineering and Applied Physics, Ginzton Laboratory, Stanford University, Stanford, CA 94305-4090 USA (e-mail: diwakar@stanfordalumni.org).

C. Debaes is with the Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Brussels 1050, Belgium.

Digital Object Identifier 10.1109/JSTQE.2003.813309

significantly reduces the delay in routing and driving signals to optical components. Kyriakis–Bitzaros *et al.* [11], on the basis of a realistic model in $0.8\ \mu\text{m}$ CMOS technology, demonstrated that the latency of an optical link with VCSELs and edge-emitting lasers is lower than the electrical link even for subcentimeter line length.

Most previous work has looked at the latency of an optical link using an NRZ data format with a VCSEL or an edge-emitting laser as a transmitter. The turn-on delay of lasers could add significant latency, which depends on the electrical drive signal strength and waveform [12]. This turn-on delay can be eliminated by using modulators instead of VCSELs. It is additionally possible to significantly reduce the latency of optical interconnect receivers by using short pulses with modulators. The fast optical rise time and concentration of all the energy in short pulses both work toward reducing the latency.

Short pulses provide many advantages in communications [13], though it is important that the medium be able to support the propagation of these pulses. On electrical wires, the frequency dependent losses are very high for short pulses, and there is substantial dispersion that spreads the pulses, making their use impractical. In optics, the losses for the entire spectrum of short pulses are nearly constant. Also, the dispersion in an optical medium for small distances is tolerable and does not cause significant broadening. It has been demonstrated that use of short pulses in modulator based interconnects can remove skew and jitter from an array of modulators [14], provide sensitivity enhancement in receivers [15], deliver a precise clock signal [17], and allow single source wavelength division multiplexed interconnects [18]. All these advantages of short pulses make it an attractive option for short distance optical interconnects.

In parallel optical interconnects, the latency is reduced by reducing complexity. Instead of multiplexing many data streams and sending them over fewer channels, multiple channels running at the clock rate of the given technology allow simpler receiver and transmitter circuits, which in turn reduce the latency of the link. Note that the issue in this paper is not one of the total bit rate that can be supported by different approaches, nor is it one of the efficiency of utilization of optical bandwidth. In longer communications links, long delay is unavoidable because of signal propagation times and receiver latencies are not important. In contrast, the receiver delay (latency) in short interconnects can be a substantial fraction of the overall delay, which in turn is very important for system performance. This paper, therefore, deals with the comparison of latencies between NRZ and short pulse links in different receiver architectures through simulation and measurement, at roughly the system clock rate supported by the given technology. The organization of this paper is as follows: the concept of a modulator-based interconnect system is described in Section II. Section III presents simulations of three different receiver topologies. Latency measurement results for the transimpedance receiver are presented in Section IV. Finally, conclusions are given in Section V.

II. SYSTEM CONCEPT

A reflection-modulator-based interconnect system is shown in Fig. 1. The optical devices on the chips are *p-i-n* diodes with quantum wells in the intrinsic region. The devices are referred

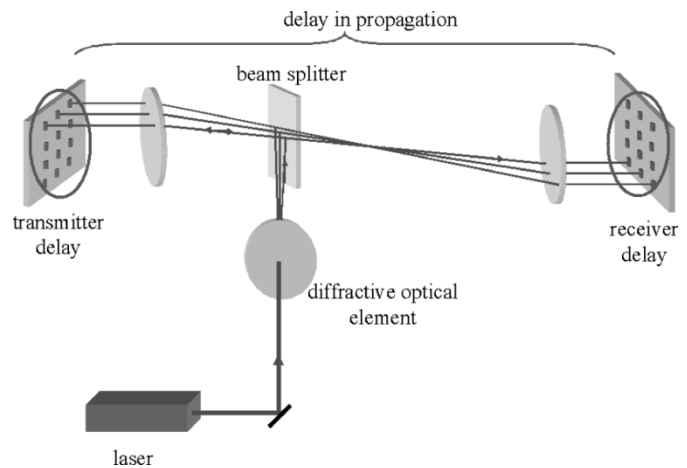


Fig. 1. Components of latency in a modulator-based interconnect system.

to here as multiple quantum well (MQW) diodes. A detailed description of the operation of these devices is given in [19]. MQW diodes can act as both the modulator and the detector, depending on the circuit to which they are connected. These diodes are hybrid-integrated to silicon CMOS chips [1] to take advantage of the highly advanced silicon CMOS process and the good optoelectronic properties of GaAs at $850\ \text{nm}$, the wavelength of operation. In a modulator-based interconnect system, a fan-out element generates multiple beams from the input beam. These beams are modulated by the reflection-modulator array and imaged on the receiver array, where the data is recovered.

Optical interconnects have three components: the transmitter, the medium of propagation and the receiver. The transmitter can be easily optimized because it essentially consists only of digital components (its input is a digital logic level). For a MQW modulator, the driver is typically an electrical buffer chain and optimization of such a circuit is outlined in [20]. The receiver, having analog input, provides the largest room for improvement [21]. Therefore, in the following sections, we analyze the latency of different receiver architectures for NRZ and short pulse inputs. Signal latency, here, is defined as the maximum delay between rise or fall of the input and output waveforms, measured at 50% of the signal amplitude.

The following parameter values are assumed for simulation, which correspond to the parameters of $0.25\ \mu\text{m}$ technology in which the receivers were fabricated. Clock period in a given technology is roughly eight FO-4, where FO-4 is the delay of an inverter driving another inverter four times its size [5]. In $0.25\ \mu\text{m}$ technology, this corresponds to about 1 GHz clock rate and the link is assumed to operate with this clock. Low-capacitance high-responsivity photodiodes are assumed based on results presented in the literature [22]. (The actual diode capacitance was higher in current fabrication, used in our experimental results.)

- Supply voltage = 2.5 V.
- Speed of operation = 1 Gb/s.
- Photodiode responsivity = 0.5 A/W.
- Number of post amplifier stages = 2.
- Capacitance of the diode after integration = 40 fF.
- Optical pulsewidth is subpicosecond.

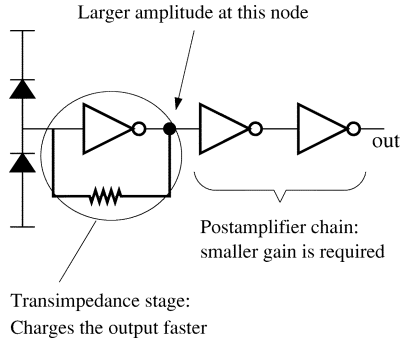


Fig. 2. Illustration of latency reduction mechanism in a transimpedance receiver with short pulse input.

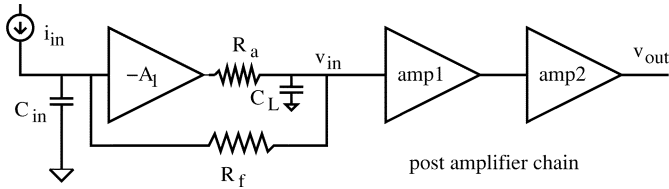


Fig. 3. First-order model of a transimpedance receiver with variable length post-amplifier chain.

- Pulswidth of electrical current pulses generated from photodiode (limited by the transit time of carriers in intrinsic region) = 10 ps.

III. MODELING OF THE RECEIVER

A. Transimpedance Receivers

The transimpedance receiver is the most commonly used receiver in optical communication. The first stage of a transimpedance receiver consists of an amplifier with resistive feedback. This is followed by gain stages. In this paper, the transimpedance receiver is based on inverters following the implementation used in [23] and [24]. Fig. 2 shows a simplified schematic of the transimpedance receiver. Intuitively, we would expect to lower the latency of the transimpedance receiver by using short pulses, instead of NRZ signaling because, for the same total energy in the bit period, a larger maximum amplitude at the output of the transimpedance stage is generated with short pulse input. This larger amplitude reduces the gain required from later stages, hence reducing the latency. This circuit was analyzed using the circuit-simulator SPICE, and a first-order analytic model. The latency of the transimpedance receiver with an NRZ data format was analyzed in [21].

To understand the mechanism of latency in receivers, a simplified model of the transimpedance receiver is analyzed. This model is shown in Fig. 3. The first stage is the transimpedance amplifier with a finite gain-bandwidth product. An ideal amplifier with a series output impedance R_a (same as $1/g_{ds}$ of transistors), together with the output capacitance, models the finite-gain-bandwidth amplifier. All the capacitances at the output of the front-end amplifier, including the input capacitance of the next stage, are combined into a single capacitance represented by C_L . The gain stages are modeled as open loop amplifiers with a finite -gain-bandwidth product. After computing the swing at the output of the transimpedance amplifier, the re-

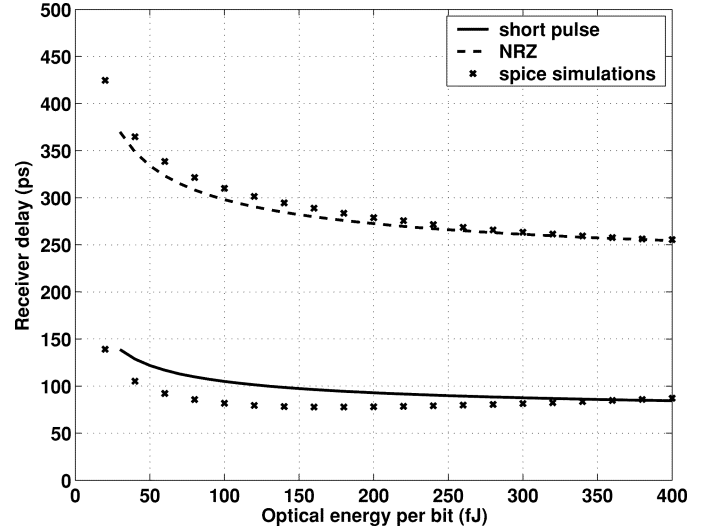


Fig. 4. Receiver delay versus optical energy per bit for short pulse and NRZ inputs, for the first-order model (lines). Corresponding SPICE simulations are denoted with "x."

quired gain-per-stage (G_{ps}) is calculated for the post-amplifier chain. Due to the finite-gain-bandwidth product, the time constant of the stage can be deduced given the required G_{ps} . A first-order estimation of latency, defined as the delay from 50% input change to 50% output change, can be performed by adding the time constants of all these stages. A step input simulates the NRZ input and a 10-ps pulse simulates the short pulse input.

The following parameter values are assumed.

- Total capacitance at the input of the receiver (C_{in}) = 90 fF.
- Feedback resistance (R_f) = 5 k Ω .
- Output impedance of the amplifier (R_a) = 8 k Ω .
- Total capacitive loading at the output of the amplifier (C_L) = 60 fF.
- Open loop gain of the amplifier (A_1) = 15.
- Gain-bandwidth product of each post amplifier stage = 10 GHz.

The transfer function of the transimpedance stage is given by

$$H(s) = \frac{V_{out}(s)}{I_{in}(s)} = \frac{R_a - A_1 R_f}{1 + A_1 + s\tau + s^2/\omega_0^2} \quad (1)$$

where $\tau = R_f C_{in} + R_a C_{in} + R_a C_L$ and $1/\omega_0^2 = R_f C_{in} R_a C_L$. Based on this transfer function, pulse and step responses were computed for the transimpedance stage. Energies per bit were computed based on 1-Gb/s operation of the receiver.

Receiver latency with different input optical energies is shown in Fig. 4. This result shows a latency reduction of $\sim 65\%$ for large optical energies by using short pulses as compared with NRZ [25]. The results match the precise SPICE simulations for the transimpedance receiver. It should be noted though, that increasing the data rate reduces the advantage of short pulses over NRZ, and in the limit when bit period is the same as the short pulsewidth, they both have equivalent performance.

The effect on latency by changing the number of post-amplifier stages is considered next. If the total gain needed from the

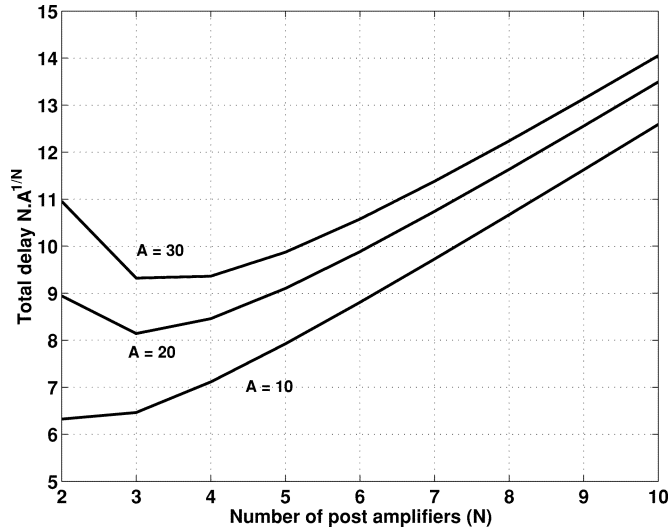


Fig. 5. Variation of receiver delay versus number of post-amplifier stages for different total gains, assuming a constant gain-bandwidth product for all stages.

post amplifier chain is A and the number of stages is N , then the gain required per stage is $G_{ps} = A^{1/N}$. Due to finite gain-bandwidth product of each stage, the delay per stage (inverse of bandwidth) is proportional to the gain required per stage. Hence the total delay of the chain (τ_d) scales as

$$\tau_d \propto NA^{1/N} \quad (2)$$

G_{ps} decreases exponentially with the number of stages N . For low N , the exponential decay of $A^{1/N}$ dominates, while for a larger N the linear increase of N dominates. This behavior can be seen in Fig. 5. Intuitively, for a large N , when N is increased to $N + 1$, the reduction in gain per stage is very small. Since the reduction in gain is small, the reduction in delay per stage is also small, but because of the extra stage the total delay (which is the sum of the delays of all stages) increases. On the other hand, for a small N , when N is increased to $N + 1$, the reduction in gain per stage is relatively large causing a large reduction in the delay. Even with one extra stage, the overall delay is reduced. This implies that there is an optimal number of post-amplifier stages to minimize latency.

The receiver delay versus the number of stages for different input optical energies per bit is shown in Fig. 6. This plot follows the same pattern as in Fig. 5. For optical energy around 100 fJ, two or three stages of post-amplifier minimizes the latency of the receiver. The calculated receiver latency versus optical energy for a two-stage and a three-stage post amplifier is shown in Fig. 7. This figure illustrates that as the pulse energy is increased, the amount of gain required reduces, causing the delay to be minimized by a lower number of stages for a pulse energy higher than a certain crossover pulse energy. Crossover occurs at ~ 70 fJ for NRZ in this figure, but for short pulses this crossover occurs below the plotted optical energies.

In this section, we saw that by using short pulses the latency in the transimpedance receiver can be reduced by more than 60% compared with using NRZ data. The results of the first order model and SPICE simulation match very closely. By using the first order model, it was also concluded that for a given energy

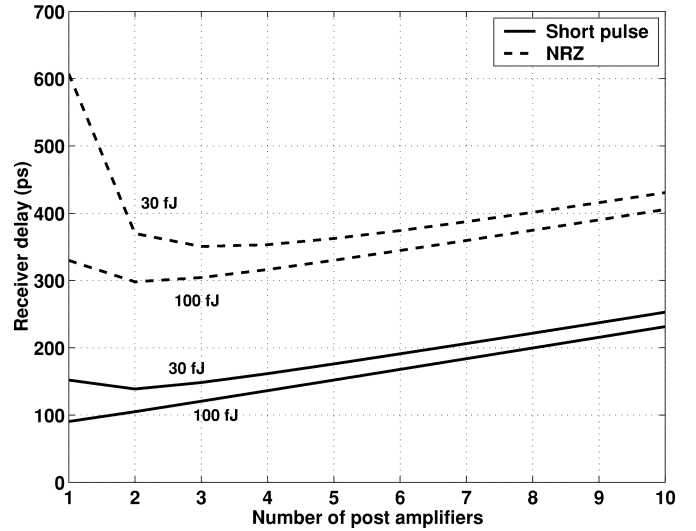


Fig. 6. Receiver delay versus number of post-amplifier stages for different optical energy per bit.

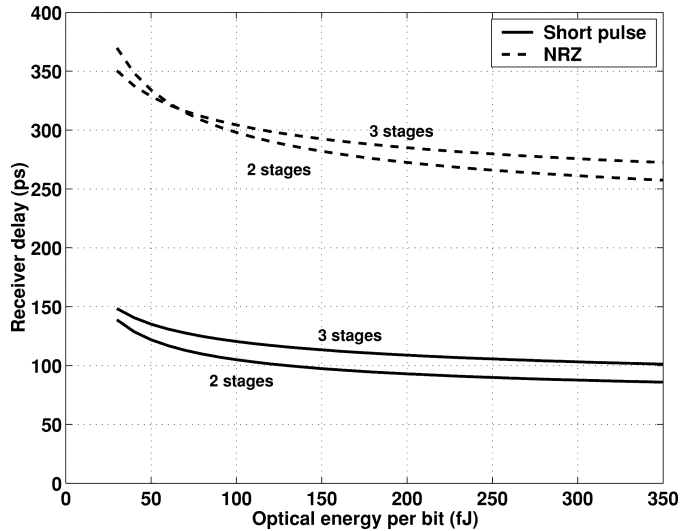


Fig. 7. Receiver delay versus optical energy per bit for 2 and 3 post-amplifier stages.

per bit, there is an optimum number of stages to minimize latency, which may not be the same for short pulses and NRZ input. For reasonable pulse energies, as a rule of thumb, the latency is minimized by using somewhere between two to five post-amplifier stages. Latency measurement results for this receiver are given in Section IV.

B. Integrating Receiver

Fig. 8 shows the circuit schematic diagram of the integrating receiver front-end. The architecture of this integrating receiver is based on a strongarm latch [16]. This receiver regeneratively amplifies the differential input to generate logic levels. It integrates the input photocurrent for half a cycle and for the remaining half cycle it evaluates based on integrated charge. The output of this front-end is valid for half the cycle and for the other half it is set to the supply voltage. A set-reset (SR) latch is used to convert this output to a valid output for the entire bit period.

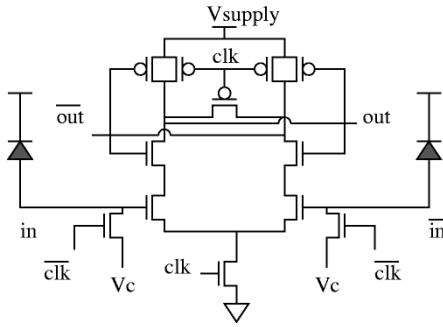


Fig. 8. Circuit schematic of the integrating receiver front-end.

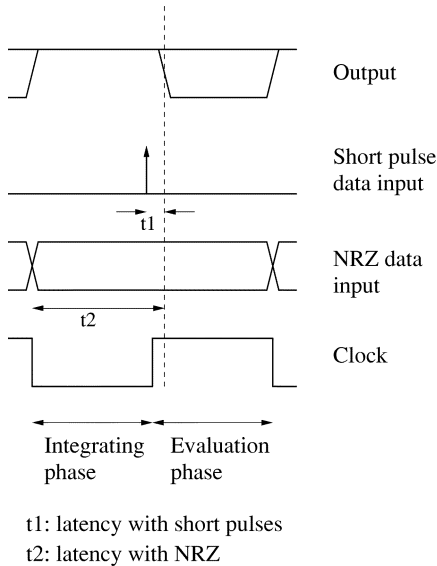


Fig. 9. Latency with respect to the clock in the integrating receiver with NRZ and short pulse inputs.

Fig. 9 illustrates the timing for the integrating receiver. In this receiver, the latency is a function of the total integrated charge. A typical integration period is half of the clock cycle. If the energy is spread over the entire bit period, as in the case of NRZ, the latency is equal to the integration period (half of the bit period) plus the time to resolve the logic level (at any bit rate). Changing the duty cycle from 50% leads to a tradeoff of energy and latency. Reducing the integration period reduces latency but increases the required optical energy per bit, while increasing the integration period has the opposite effect. In the case of short pulses, the pulses can arrive at the end of the integration period and transfer all of the energy in an instant. As seen in the figure, the latency with a short pulse can be as low as the evaluation time. For a practical system, there needs to be some timing margin to account for the jitter and other variability in the system.

This receiver operates on the principle of positive feedback, hence it is very sensitive. The evaluation time of this receiver depends logarithmically on the amount of integrated charge. For modeling this receiver, the parameter values of the $0.25\ \mu\text{m}$ CMOS technology were assumed so that the results could be compared with the transimpedance receiver. A SPICE simulation of the latency of the entire integrating receiver circuit (including the SR latch) is plotted in Fig. 10. According to these

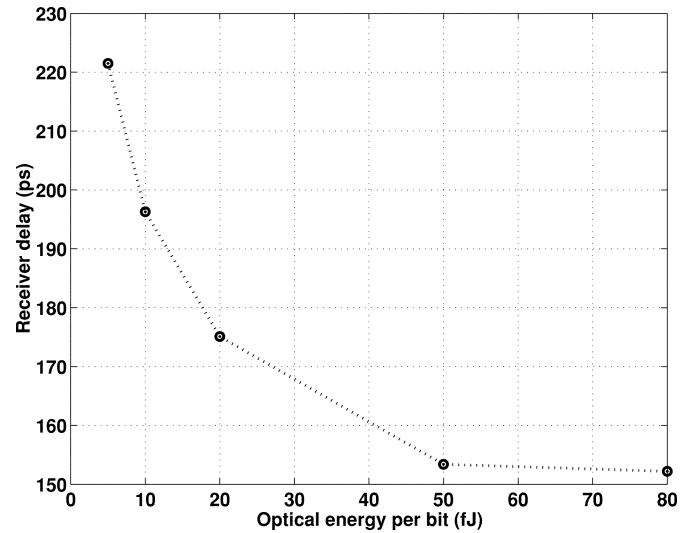
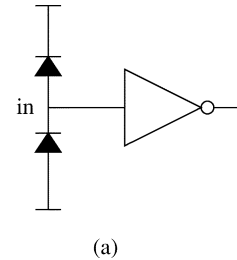
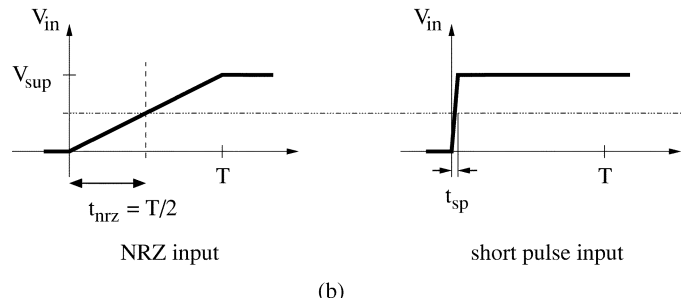


Fig. 10. Latency of the entire integrating receiver, including the SR latch, with short pulse input computed by using SPICE circuit simulator.



(a)



(b)

Fig. 11. (a) Schematic of the totem-pole diode pair receiver connected to the high impedance input of an inverter buffer. (b) Timing diagram at node "in" with NRZ and short pulse input.

results, the total latency of the receiver is ~ 150 ps for 50 pJ of pulse energy. At 1 Gb/s operation, the latency of the receiver is ~ 500 ps + 150 ps = 650 ps with NRZ input, while it is 150 ps for short pulse input as shown above. Thus, the latency of this receiver is also significantly reduced by using short pulses.

C. Totem-Pole Diode Receiver

Very low latency at the expense of larger optical power can be achieved by using a diode pair connected in the totem-pole configuration as shown in Fig. 11. This design is effectively receiver-less ("reless") as there is no voltage amplifier involved. Removing amplifier stages can not only reduce the delay of the receiver but also eliminate skew and jitter introduced by these stages. By reducing skew and jitter, a very precise optical clock can be injected with short pulses

as mentioned in [17]. Elimination of amplifiers also allows to deliver very fast edges through this “receiver,” which can be utilized for characterizing high speed signals on-chip.

This “receiver” needs to be connected to a high impedance, such as the input of a buffer, where the charge is integrated. Here the input capacitance is charged to the supply rails by providing sufficient optical energy. The optical energy required to charge the node “in” to the supply rails is a linear function of the front-end capacitance, which is typically dominated by the photodiode capacitance. If the total capacitance at the node “in” is C_{in} and the total voltage swing required is V_{sup} , then the total charge required is $Q_{tot} = C_{in}V_{sup}$. For a photodiode responsivity R , the minimum optical energy required is $E_{opt} = C_{in}V_{sup}/R$. This optical energy can either be delivered in a very brief period or it can be spread out over the entire bit period (T). If the input to this receiver is NRZ data with the minimum required pulse energy, the input node will reach half of the supply voltage in half a cycle ($t_{nrz} = T$). If short pulses are used, instead, the input node will be charged quickly (t_{sp}), limited only by the carrier transit time in the intrinsic region of the diode. The timing diagram in Fig. 11 shows the charging of the input node with NRZ and short pulses.

If the flip-chip bonded photodiode capacitance is 40 fF and the responsivity is 0.5 A/W, then for a total capacitance of 90 fF (i.e., assuming a buffer capacitance of ~ 10 fF) the optical energy required to charge the input node by 2.5 V (i.e., the supply voltage for 0.25 μm CMOS technology) is 450 fJ. At 1 Gb/s this translates to 450 μW of optical power. This is a relatively large amount of power, which may make practical use of this receiver difficult. In specific applications, though, this might be an acceptable power level. By using a metal–semiconductor–metal photodiode, or a silicon photodiode in a silicon-on-insulator process, the photodiode capacitance can, however, be reduced, which can reduce the optical energy required. For a 1- μm long intrinsic region, the carrier transit time is roughly 10 ps, which determines the latency with short pulses in this receiver. By comparison, for 1 Gb/s operation, the latency with NRZ data and minimum optical power will be 500 ps. Hence, this receiver gives the minimum latency with short pulses of all three receivers studied here, though the amount of optical energy required can be much larger.

IV. MEASUREMENT OF THE TRANSIMPEDANCE RECEIVER LATENCY

The results in the earlier section predict that receiver latency can be significantly reduced by using short pulses. To verify this concept, the latency of the receiver–modulator driver pair was measured experimentally. Circuits were fabricated in 0.25- μm standard CMOS technology, and the optical devices, MQW diodes, were flip-chip bonded on these circuits. The active area of these diodes was 20 $\mu\text{m} \times 20 \mu\text{m}$. The schematic of the transmitter–receiver circuit is shown in Fig. 12. A PMOS transistor act as the resistive feedback element. By varying the voltage at node “tune,” the resistance value can be changed. Node “tune” was kept at 0 V for latency measurement. An optical pump-probe setup was used for measurement, which is

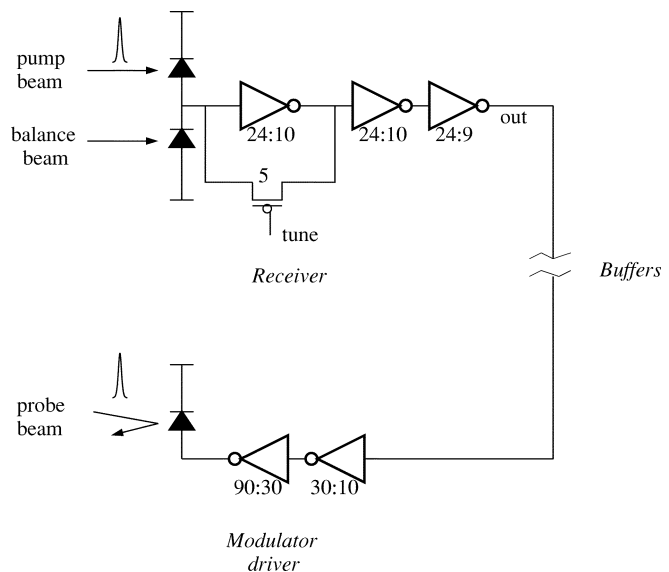


Fig. 12. Receiver-transmitter module used for testing latency via the pump-probe method. The numbers shown here are the sizes of PMOS and NMOS transistors in λ , where $\lambda = 0.2 \mu\text{m}$.

described in detail in [26]. Short pulses (~ 150 fs) generated from a Ti : sapphire modelocked laser at 850 nm were used for the pump and probe beams at 80 MHz (limited by the repetition rate of the laser). *Pump beam* and a cw laser output as *balance beam* were incident on the differential diode pair at the receiver input. The pump beam excited the receiver, while the balance beam brought the receiver back to its original state over time. A modulator driver was driven by the output of the receiver. The voltage output of the modulator driver was sampled optically with a *probe beam* at the same rate as the *pump beam*. Varying the delay between pump and probe beam mapped the response of the transceiver pair. Since the optical pulses were only 150 fs, subpicosecond resolution could be achieved in measurements.

The measurement of latency for NRZ data was performed on a different setup, because the delays were much larger. They were done using a high speed detector (2.5-GHz bandwidth), and directly evaluating the waveforms on an oscilloscope. The receiver was designed to operate at 1 Gb/s, but since the repetition rate of the short pulse laser was 80 MHz, all the measurements were done at 80 Mb/s. Measurement results were extrapolated to 1 Gb/s for both NRZ and short pulses. For a given receiver design this does not change the latency, which is the primary concern of this paper. Additionally, the optical energy per bit reduces proportionally with the speed of operation for NRZ data for a given latency, while it remains constant for short pulses. This implies that increasing the bit rate works in favor of NRZ. In the limit of the bit period equal to the short pulsewidth, both data formats are equivalent. Here, we assume the operation of the link is at the system clock rate supported by 0.25 μm process, i.e., 1 Gb/s. The latency of an entire interconnect can be easily computed by adding the delay of signal propagation to the measured latency of the transceiver pair.

Fig. 13 shows the measured values of latency for NRZ and short pulses for the circuit in Fig. 12. These results match the SPICE simulations of the circuit reasonably well. It is evident

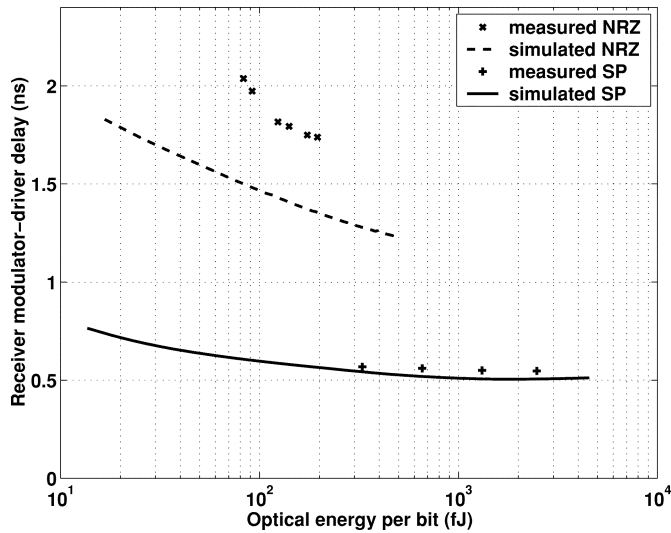


Fig. 13. Comparison of the latency of the transimpedance receiver-transmitter module with short pulse and NRZ inputs.

that short pulses reduce the latency of the transceiver pair compared with NRZ input by a very significant amount, as predicted by the simplified modeling in Section III. In these measurements, the diode capacitance was ~ 260 fF though the circuits were designed to drive lower capacitance. This capacitance can be easily reduced to less than 50 fF by improved fabrication techniques and using smaller devices. Lower capacitance and optimized circuit will reduce the latency further.

Measurements of latency of a transimpedance receiver implemented in bipolar technology were presented by Wieland *et al.* [27]. The overall delay of their receiver was 1.5 ns at 1-Gb/s operation. The latency measured with NRZ data here for the transceiver is of the same order, though the latency is much lower with short pulses.

V. CONCLUSIONS AND DISCUSSION

Chip sizes are expected to increase modestly with future generations, and will remain around 2 cm on a side. Assuming a global interconnect of 2 cm the latency of a repeatered electrical line is ~ 330 ps (at 20% of the velocity of light) [4]. For optical interconnects, the propagation time for a 2 cm distance in glass is ~ 100 ps. The latency in the transmitter can be brought down to ~ 70 ps (assuming a single buffer driving the modulator capacitance), and as we have shown in Section III, the transimpedance receiver latency can be reduced to ~ 70 ps with short pulses. It is even possible to have less than 70 ps of receiver latency by using the “recless” receiver. Thus, this shows that optical interconnects can achieve latencies comparable to electrical interconnects or even less, at least theoretically, for on-chip global communication. The key reasons for low latency in optical interconnects are advanced integration of optical devices with silicon, and short pulse encoding.

In this paper, an analysis of receiver latency for different architectures was presented. It is shown that short pulse encoding significantly reduces the receiver latency compared with NRZ

TABLE I
RECEIVER LATENCY WITH NRZ AND SHORT PULSE INPUTS

Receiver type	NRZ delay (ps)	short pulse delay (ps)
Transimpedance	340	120
Integrating	650	150
Recless	500	10

encoding. Table I summarizes the results of the receiver latency simulation. A “recless” receiver with short pulses has the shortest delay, but at the expense of optical power. The optical power required depends on the photodiode capacitance. The transimpedance and integrating receivers have nearly the same latency with short pulses.

The effects of various parameters on the latency of the transimpedance receiver by using a simplified model were described. It was shown that the number of stages required to minimize the latency depends on the input optical energy, i.e., as the energy per bit is increased, the optimal number of stages reduces. Finally, experimental measurements on a transimpedance receiver demonstrate latency reduction by more than 50% using short pulses, verifying modeling and simulation. With the same optical energy per bit, short pulses substantially reduce the latency for a given transimpedance design compared to NRZ.

This work is an initial study of the use of short optical pulses to reduce latency in optical interconnects. Short pulses hold a vast potential in optical interconnects. The availability of compact high-speed mode-locked lasers, packaging, and the cost are the current challenges in making a viable intrachip and interchip optical interconnect system. With further advances in optical devices and integration technology it might be possible to have low latency optical interconnects using short pulses for on-chip communication.

ACKNOWLEDGMENT

The authors would like to thank Dr. P. Atanackovic for MBE growth of the GaAs material and Dr. R. Piestun for the design of diffractive optical element. National Semiconductor Corporation provided silicon CMOS chip fabrication. D. Agarwal would also like to acknowledge discussions with Dr. M. Horowitz and Dr. J. W. Goodman.

REFERENCES

- [1] K. Goossen, J. Walker, L. D’Asaro, S. Hui, B. Tseng, R. Leibenguth, D. Kossives, D. Bacon, D. Dahringer, L. Chirovsky, A. Lentine, and D. Miller, “GaAs MQW modulators integrated with silicon CMOS,” *IEEE Photon. Technol. Lett.*, vol. 7, pp. 360–362, Apr. 1995.
- [2] R. Pu, C. Duan, and C. Wilmsen, “Hybrid integration of VCSEL’s to CMOS integrated circuits,” *IEEE J. Select. Topics Quantum Electron.*, vol. 5, pp. 201–208, Mar. 1999.
- [3] “The International Technology Roadmap for Semiconductors (2001 Edition),”

- [4] D. Miller, "Rationale and challenges for optical interconnects to electronic chips," *Proc. IEEE*, vol. 88, pp. 728–749, June 2000.
- [5] M. Horowitz, C. Yang, and S. Sidiropoulos, "High-speed electrical signaling: Overview and limitations," *IEEE Micro*, vol. 18, pp. 12–24, Jan. 1998.
- [6] A. Deutsch, P. Coteus, G. Kopsay, H. Smith, C. Surovic, B. Krauter, D. Edelstein, and P. Restle, "On-chip wiring design challenges for gigahertz operation," *Proc. IEEE*, vol. 89, pp. 529–555, Apr. 2001.
- [7] A. Krishnamoorthy and D. Miller, "Scaling optoelectronic-VLSI circuits into the 21st century: A technology roadmap," *IEEE J. Select. Topics Quantum Electron.*, vol. 2, pp. 55–76, Apr. 1996.
- [8] J. Dambre, H. Van Marck, and J. Van Campenhout, "Quantifying the impact of optical interconnect latency on the performance of optoelectronic FPGAs," in *Proc. 6th Int. Conf. Parallel Interconnects*, Oct. 1999, pp. 91–97.
- [9] J. Collet, D. Litaize, J. VanCampenhout, C. Jesshope, M. Desmulliez, H. Thienpont, J. Goodman, and A. Louri, "Architectural approach to the role of optics in monoprocessor and multiprocessor machines," *Appl. Opt.*, vol. 39, pp. 671–682, Feb. 2000.
- [10] H. Neefs, P. Van Heuven, and J. Van Campenhout, "Latency requirements of optical interconnects at different memory hierarchy levels of a computer system," in *Proc. SPIE Optics Computing*, vol. 3490, 1998, pp. 552–555.
- [11] E. Kyriakis-Bitaros, N. Haralabidis, Y. Moisiadis, M. Lagadas, A. Georgakilas, and G. Halkias, "Comparison of the signal latency in optical and electrical interconnections for interchip links," *Opt. Eng.*, vol. 40, pp. 144–146, Jan. 2001.
- [12] D. Cutrer and K. 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, Jan. 1995.
- [13] K. Tamura, "Short pulse lasers and their applications to optical communications," in *Proc. 12th Annu. Meeting IEEE Lasers and Electro-Optics Society*, vol. 2, 1999, pp. 537–538.
- [14] G. Keeler, B. Nelson, D. Agarwal, and D. Miller, "Skew and jitter removal using short optical pulses for optical interconnection," *IEEE Photon. Technol. Lett.*, vol. 12, pp. 1041–1135, June 2000.
- [15] G. Keeler, D. Agarwal, B. Nelson, N. Helman, and D. Miller, "Performance enhancement of an optical interconnect using short pulses from a modelocked diode laser," in *Proc. Conf. Lasers and Electro-Optic Society*, May 2002, p. 163.
- [16] J. Montanaro *et al.*, "A 160-MHz, 32-b, 0.5-W CMOS RISC microprocessor," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1703–1714, Nov. 1996.
- [17] C. Debaes, D. Agarwal, A. Bhatnagar, H. Thienpont, and D. Miller, "High-impedance high-frequency silicon detector response for precise receiverless optical clock injection," in *Proc. SPIE*, vol. 4654, 2002, pp. 78–88.
- [18] B. Nelson, G. Keeler, D. Agarwal, N. Helman, and D. Miller, "Demonstration of a wavelength division multiplexed chip-to-chip optical interconnect," in *Proc. Conf. Lasers Electro-Optic Society*, May 2002, pp. 163–164.
- [19] G. Boyd, D. Miller, D. Chemla, S. McCall, A. Gossard, and J. English, "Multiple quantum well reflection modulator," *Appl. Phys. Lett.*, vol. 50, pp. 1119–1121, Apr. 1987.
- [20] B. Cherkauer and E. Friedman, "A unified design methodology for CMOS tapered buffers," *IEEE Trans. VLSI Syst.*, vol. 3, pp. 99–111, Mar. 1995.
- [21] E. Kyriakis-Bitaros, N. Haralabidis, M. Lagadas, A. Georgakilas, Y. Moisiadis, and G. Halkias, "Realistic end-to-end simulation of the optoelectronic links and comparison with the electrical interconnections for system-on-chip applications," *J. Lightwave Technol.*, vol. 19, pp. 1532–1542, Oct. 2001.
- [22] A. Krishnamoorthy, T. Woodward, R. Novotny, K. Goossen, J. Walker, A. Lentine, L. D'Asaro, S. Hui, B. Tseng, R. Leibenguth, D. Kossives, D. Dahringer, L. Chirovsky, G. Aplin, R. Rozier, F. Kiamilev, and D. Miller, "Ring oscillators with optical and electrical readout based on hybrid GaAs MQW modulators bonded to 0.8 um silicon VLSI circuits," *Electron. Lett.*, vol. 31, pp. 1917–1918, Oct. 1995.
- [23] T. Woodward, A. Krishnamoorthy, A. Lentine, and L. Chirovsky, "Optical receivers for optoelectronic VLSI," *IEEE J. Select. Topics Quantum Electron.*, vol. 2, pp. 106–116, Apr. 1996.
- [24] M. Forbes, "Electronic design issues in high-bandwidth parallel optical interfaces to VLSI circuits," Ph.D. dissertation, Heriot-Watt Univ., Edinburgh, U.K., Mar. 1999.
- [25] D. Agarwal and D. Miller, "Latency in short pulse based optical interconnects," in *PROC. 14th Annual Meeting IEEE Lasers and Electro-Optics Society*, vol. 2, 2001, pp. 812–813.

- [26] G. Keeler, D. Agarwal, C. Debaes, B. Nelson, C. Helman, H. Thienpont, and D. Miller, "Optical pump-probe measurements of the latency of silicon CMOS optical interconnects," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 1214–1216, Aug. 2002.
- [27] J. Weiland, H. Melchior, M. Kearley, C. Morris, A. Moseley, M. Goodwin, and R. Goodfellow, "Optical receiver array in silicon bipolar technology with selfaligned, low parasitic III/V detectors for DC-1 Gbit/s parallel links," *Electron. Lett.*, vol. 27, pp. 2211–2213, Nov. 1991.



Diwakar Agarwal (S'95) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, India, in 1994 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1996 and 2002, respectively. His Ph.D. research was done under Prof. David A. B. Miller.

His research interests involve optical–electrical electrical–optical interfaces, and optical interconnect systems.



Gordon A. Keeler (S'97) was born in Thunder Bay, Canada, in 1973. He received the H.B.Sc. degree in physics from Lakehead University, Thunder Bay, Canada, in 1996 and the M.S. and Ph.D. degrees in applied physics from Stanford University, Stanford, CA, in 1998 and 2003, respectively. His Ph.D. research was done under Prof. David A. B. Miller.

His research involved optical interconnects and optoelectronic devices.



Christof Debaes (S'99) was born in Geraardsbergen, Belgium, in 1975. He graduated as an electrotechnical engineer from the Vrije Universiteit Brussel (VUB) in 1998. He received the Ph.D. degree from the Applied Physics and Photonics Department, VUB, in collaboration with the Ginzton Laboratory, Stanford University, Stanford, CA, directed by D. A. B. Miller.

He has been granted a postdoctoral fellowship from the Flemish Fund for Scientific Research.

His research interests involve short-distance optical interconnects and clock distribution networks to CMOS-chips.



Bianca E. Nelson (S'97) was born in Milwaukee, WI, in 1973. She received the B.S. degree in physics and astronomy from the University of Iowa, Iowa City, in 1996 and the M.S. and Ph.D. degrees in applied physics from Stanford University, Stanford, CA, in 1998 and 2003, respectively. Her Ph.D. research was done under Prof. David A. B. Miller.

Her research involved wavelength division multiplexed optical interconnect systems and passive optical devices.



Noah C. Helman (S'02) received the A.B. degree in physics from Harvard University, Cambridge, MA, in 1998 and the M.S. degree in applied physics from Stanford University, Stanford, CA, in 2002. He is currently working toward the Ph.D. degree in applied physics under Prof. David A. B. Miller at Stanford University.

His research involves semiconductor modulators in optical interconnect systems.



David A. B. Miller (M'84-SM'89-F'95) received the B.Sc. degree from St. Andrews University, St. Andrews, U.K. and the Ph.D. degree from Heriot-Watt University, Edinburgh, U.K., in 1979.

He was with Bell Laboratories, from 1981 to 1996, as a department head from 1987, latterly of the Advanced Photonics Research Department. He is currently the W. M. Keck Professor of Electrical Engineering at Stanford University, Stanford, CA, and the Director of the Ginzton and Solid State and Photonics Laboratories, Stanford, CA. His research interests include quantum-well optoelectronic physics and devices, and fundamentals and applications of optics in information, sensing, switching, and processing. He has published more than 200 scientific papers, holds over 40 patents, and has received several awards for his work.

Dr. Miller has served as a Board member for both the Optical Society of America (OSA) and IEEE Lasers and Electro-Optics Society (LEOS), and in various other society and conference committees. He was President of the IEEE Lasers and Electro-Optics Society in 1995. He was awarded the Adolph Lomb Medal and the R. W. Wood Prize from the OSA, the International Prize in Optics from the International Commission for Optics, and the IEEE Third Millennium Medal. He is a Fellow of the Royal Societies of London and Edinburgh, OSA, and APS, and holds an honorary degree from the Vrije Universiteit Brussels, Brussels, Belgium.

Dr. Miller has served as a Board member for both the Optical Society of America (OSA) and IEEE Lasers and Electro-Optics Society (LEOS), and in various other society and conference committees. He was President of the IEEE Lasers and Electro-Optics Society in 1995. He was awarded the Adolph Lomb Medal and the R. W. Wood Prize from the OSA, the International Prize in Optics from the International Commission for Optics, and the IEEE Third Millennium Medal. He is a Fellow of the Royal Societies of London and Edinburgh, OSA, and APS, and holds an honorary degree from the Vrije Universiteit Brussels, Brussels, Belgium.