

Optical Pump-Probe Measurements of the Latency of Silicon CMOS Optical Interconnects

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

Abstract—We present the first measurements of optical-electrical-optical conversion latency in a hybridly-integrated optoelectronic/silicon complementary metal-oxide-semiconductor (CMOS) chip designed for optical interconnection. Using an optical pump-probe technique, we perform precise measurements with picosecond resolution that closely match our simulations. Our findings suggest that optical interconnects have the potential to provide equal or lower latency than on-chip global wires in future CMOS microelectronics.

Index Terms—CMOS integrated circuits, delay estimation, optical interconnections, ultrafast optics.

I. INTRODUCTION

OPTICAL interconnects may have the ability to alleviate the looming complementary metal-oxide-semiconductor (CMOS) interconnect bottleneck by providing extremely dense, high-bandwidth links in chip-to-chip and, potentially, on-chip applications. While many of the advantages of an optical approach have been demonstrated experimentally, an important parameter that has not yet been properly characterized is interconnect delay, or the signal latency of optical interconnects.

The latency of electrical interconnects is already sufficiently high that data on many global wires (e.g., the long interconnections between functional blocks) cannot be transmitted across a modern microprocessor within a single clock cycle. Using realistic scaling assumptions, it is likely that signal propagation velocity on the best repeatered resistive capacitive (RC) lines will actually decrease in the future [1]. When this is combined with larger chips and shorter clock periods, the transit time for optimized global interconnects may approach ten or more clock cycles. As interconnect delay increases, so do the problems of high power dissipation, increased design complexity, and timing uncertainty. Low-latency interconnections permit larger areas of synchronicity, thereby easing circuit design and requiring fewer power-intensive latches. Most off-chip applications, including communicating within multiprocessor systems and accessing memory elements, also benefit when delays are minimized. Thus, low-latency interconnects are very

Manuscript received February 11, 2002; revised March 29, 2002. This work was supported by the Defense Advanced Research Projects Agency under the University Optocenters program (Contract MDA972-00-1-0024) through the University of New Mexico, and by DARPA/MARCO under the Interconnect Focus Center through the Georgia Institute of Technology.

G. A. Keeler, D. Agarwal, B. E. Nelson, N. C. Helman, and D. A. B. Miller are with the Departments of Applied Physics and Electrical Engineering, Ginzton Laboratory, Stanford University, Stanford, CA 94305-4090 USA.

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

Publisher Item Identifier S 1041-1135(02)06010-X.

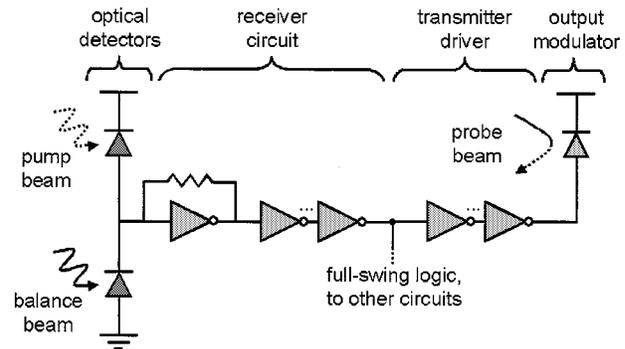


Fig. 1. Simplified diagram of the transimpedance-amplifier-based receiver and transmitter driver circuits used in the latency experiments and simulations.

desirable for optimizing future silicon microelectronics. If optics is to displace electrical wiring in chip-to-chip or on-chip links, the latency of optical interconnects must be at least comparable to, or preferably lower than, that of the electrical interconnects they replace. In this work, we perform the first measurements of optical interconnect latency, and find that these values closely match our simulation results. The delays are low enough that they would be insignificant in most off-chip applications. Furthermore, we show that the latency obtainable by a theoretical on-chip optical link using our circuits and devices would be similar to the delay of present on-chip global electrical interconnects using the same CMOS technology.

II. EXPERIMENTAL DETAILS

The signal delay of an optical link can be separated into three components: transmitter latency, time-of-flight latency, and receiver latency. Since the time-of-flight delay is simple to calculate and entirely system dependent, we are most interested in measuring and simulating the electrical-to-optical (transmitter) and optical-to-electrical (receiver) constituents. Due to the inherent speed limitations of silicon CMOS and the capacitive loading effects of electrical probing, precise measurement of these short delays is difficult using electronics. Thus, we use an optical technique and consider the two delays together, performing a measurement of the optical-electrical-optical (O/E/O) conversion latency using an electrically back-to-back receiver/transmitter pair. A simplified schematic of the O/E/O circuit used for latency measurements is shown in Fig. 1.

The circuits used in this experiment were fabricated in 0.25- μm silicon CMOS using transimpedance-amplifier-based receivers that are similar to previous designs [2]. The receivers and transmitter drivers were first tested independently; thus,

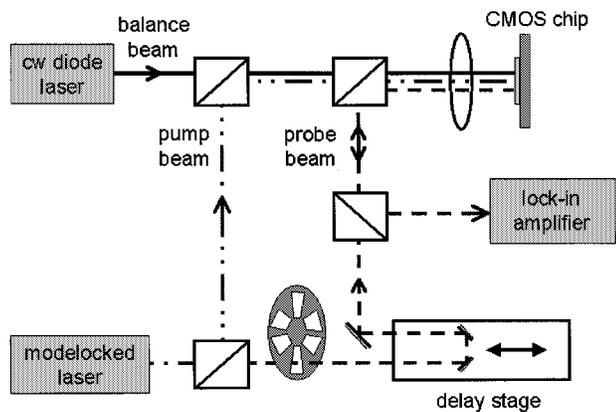


Fig. 2. Schematic of the optical pump-probe setup used for interconnect latency measurements.

the functionality of all circuits was verified at high speeds, with the receivers exhibiting full-swing logic level outputs. Both circuits are inverter-based, optically differential, and designed for gigabit-per-second operation with small area and low power consumption. To obtain high-speed and high-channel density, arrays of 200 optoelectronic I/O devices were integrated to the CMOS via flip-chip bonding. The devices are molecular beam epitaxy (MBE) grown GaAs–AlGaAs multiple-quantum-well (MQW) p-i-n diodes that serve both as modulators and as photodiodes. These electroabsorption devices, which rely upon the quantum-confined Stark effect, are used at a wavelength of 850 nm and have a capacitance of about 260 fF following integration.

The experiment was performed in an optical pump-probe setup, using the ~ 100 femtosecond pulses of a modelocked Ti:sapphire laser. Fig. 2 shows a conceptual layout. The pseudodifferential input data was comprised of a repetitive optical pulse train (pump beam) and a continuous-wave diode laser beam (balance beam), both incident on the detector pair of a differential receiver on the CMOS chip. The output voltage of this O/E/O circuit was another pulse train, whose pulses were broadened by the limited bandwidth of the circuit and time-delayed by the conversion latency. The voltage applied by the transmitter driver directly controlled the reflectivity of the modulator, which was optically sampled with a pulsed readout beam (probe beam). The intensity of the modulated probe beam could then be measured using a lock-in amplifier. By varying the relative delay between the pump and probe, the temporal response of the optoelectronic circuit was mapped with picosecond resolution.

The zero-point time, or moment at which the pump beam was incident on the input detector, was determined by performing a two-beam excitonic absorption saturation experiment. Both pump and probe beams were spatially overlapped on a single MQW modulator, and their relative delay was varied. When both pump and probe were temporally coincident, the modulator absorption became saturated, changing the amplitude of the reflected probe. Because this excitonic absorption saturation occurs on a subpicosecond time scale [3], an extremely precise zero-point time could be obtained and applied to subsequent measurements.

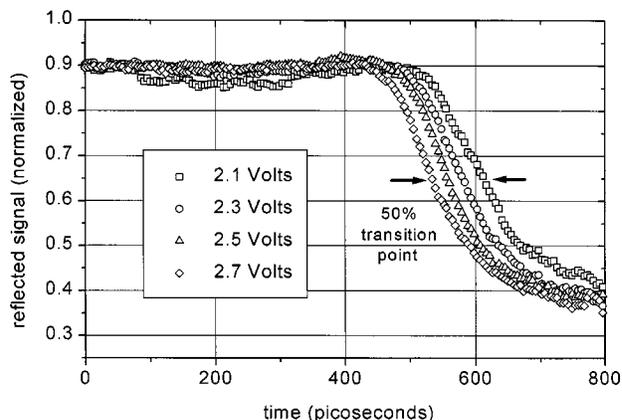


Fig. 3. Normalized data from a series of pump-probe O/E/O latency measurements when the supply voltage is varied.

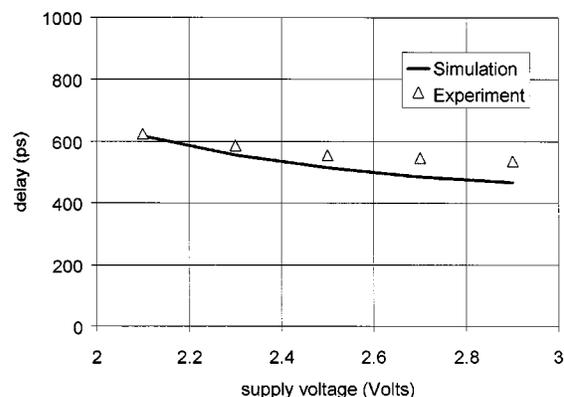


Fig. 4. Simulated and measured O/E/O latency as a function of supply voltage with a received energy of 1500 fJ/bit.

III. RESULTS

In this investigation, we define latency to be the delay between the zero-point time and the 50% transition point of the output. After measuring the capacitance and responsivity of the integrated optoelectronic devices, the latency of the transmitter and receiver circuit combination was simulated using simulation program with integrated circuit emphasis (SPICE). In our model, optical pulses from the mode-locked laser were assumed to create electrical current pulses with a 10-ps duration at the photodiode output (i.e., the input of the receiver). The response of the output modulator was assumed to be instantaneous with a change in bias voltage.

As our simulations predict, the O/E/O conversion latency was measured to range from about 500–650 ps. Fig. 3 shows the normalized results from a single receiver/transmitter pair for different supply voltages. From these curves we obtain the latency and compare it to circuit simulation results. Fig. 4 shows the simulated and measured latency as a function of supply voltage at a constant optical power. The observed variation in delay with supply voltage is important, since on-chip switching noise can cause similar fluctuations in the supply that translate to jitter at the receiver output. In Fig. 5, we compare simulations and experimental results for delay versus received energy per bit. Lower received energy corresponds to a larger latency because

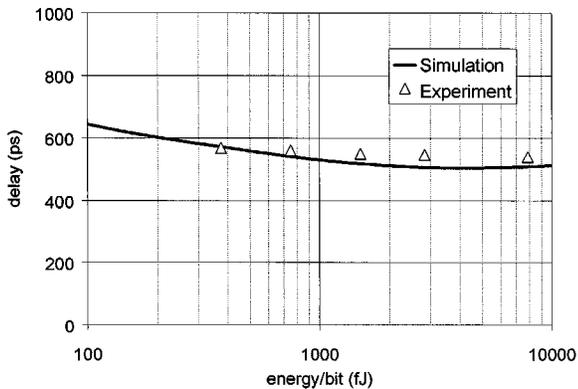


Fig. 5. Simulated and measured O/E/O latency as a function of received optical energy/bit with a 2.5-bit supply.

of the expected increase in receiver response time. The data closely matches our circuit simulation results in both cases.

IV. DISCUSSION

A repeated RC global on-chip electrical interconnect, fabricated in a 0.25- μm CMOS process, has a propagation velocity of only about 0.12 c , where c is the velocity of light in free space [4]. Thus, for a 30-mm on-chip line (i.e., roughly the edge dimension of a typical CMOS processor), such a wire will have a latency of about 800 ps. Using a recently demonstrated on-chip optical approach [5], a hypothetical optical link on the same chip would have an effective propagation velocity of roughly 0.33 c (which accounts for the refractive index of the transmission medium and a doubling of the path length because of the three-dimensional optical assembly). The resulting optical link latency is then the 300 ps required for signal propagation, plus the E/O/E conversion latency (~ 500 ps as measured here). Thus, even with our rather high-capacitance devices, an optimized global wire and an on-chip optical interconnect implemented in currently available technologies would have the same latency at a link length of 30 mm. Using lower-capacitance (~ 40 fF) integrated detectors, our simulations show that the E/O/E latency component would be reduced to around 100 ps, significantly reducing the distance at which the optical and electrical signal delays are equal.

Besides changing the optical path length, the most effective method of reducing optical interconnect latency is to minimize the receiver delay. Transmitter circuits are relatively simple and contribute little to the overall latency. Using modulators or pre-biasing vertical-cavity surface-emitting lasers (VCSELs) avoids the significant turn-on delay associated with driving a laser from zero bias. Receiver delays must be reduced through better receiver design, but will steadily improve as transistor speeds increase. It has been shown through modeling that receiver latency can be further reduced using short pulse signaling instead of nonreturn-to-zero (NRZ) signaling [6]. This latency reduction—included implicitly in our results—is just one of many benefits expected from short-pulse-based interconnects [7], [8]. Alternatively, the so-called “receiverless” detection approach [9] can be used to remove essentially all receiver delay. Such optical interconnect latency reduction techniques

can be demonstrated and analyzed using the measurement technique described here, and should have an impact on future silicon electronics by reducing the design complexity, timing uncertainty, and power dissipation associated with interconnect latency.

V. CONCLUSION

We have performed the first pump-probe measurements of optical interconnect latency on a hybridly integrated optoelectronic/CMOS chip. This technique allows optical measurements of high-speed CMOS circuit delays with a precision of a few picoseconds. The results closely match circuit simulations, confirming the validity of our technique. Comparing the measured values to the latency of current electrical interconnects, our findings demonstrate that an optical approach can already provide comparable signal delays at the global length scale. Predictions based on future latency reduction methods suggest that optical interconnects have the potential to provide lower latency than their electrical counterparts in many on-chip applications. Understanding the sources of latency and measuring actual interconnect delays with a technique such as the one presented here are the first steps toward minimizing latency for future optically interconnected silicon CMOS chips.

ACKNOWLEDGMENT

The authors wish to thank Dr. P. Atanackovic for performing the MBE wafer growth and National Semiconductor Corporation for CMOS chip fabrication. N. Helman gratefully acknowledges the support of a Gerhard Casper Stanford Graduate Fellowship, and C. Debaes is indebted for his research fellowship from the FWO Vlaanderen.

REFERENCES

- [1] P. Kapur, J. P. McVittie, and K. C. Saraswat, “Realistic copper interconnect performance with technological constraints,” in *Proc. International Interconnect Technology Conf.*, 2001, pp. 233–235.
- [2] T. K. Woodward, A. V. Krishnamoorthy, A. L. Lentine, K. W. Goossen, J. A. Walker, J. E. Cunningham, W. Y. Jan, L. A. D’Asaro, L. M. F. Chirovsky, S. P. Hui, B. Tseng, D. Kossives, D. Dahringer, and R. E. Leibenguth, “1-Gb/s two-beam transimpedance smart-pixel optical receivers made from hybrid GaAs MQW modulators bonded to 0.8- μm silicon CMOS,” *IEEE Photon. Technol. Lett.*, vol. 8, pp. 422–424, 1996.
- [3] W. H. Knox, R. L. Fork, M. C. Downer, D. A. B. Miller, D. S. Chemla, C. V. Shank, A. C. Gossard, and W. Wiegmann, “Femtosecond dynamics of resonantly excited excitons in room-temperature GaAs quantum wells,” *Phys. Rev. Lett.*, vol. 54, pp. 1306–1309, 1985.
- [4] D. A. B. Miller, “Rationale and challenges for optical interconnects to electronic chips,” *Proc. IEEE*, vol. 88, pp. 728–749, 2000.
- [5] C. Debaes, M. Vervaeke, V. Baukens, W. Meeus, P. Tuteleers, M. Brunfaut, J. Van Campenhout, and H. Thienpont, “Multi-channel free-space intra-chip optical interconnections: Combining plastic micro-optical modules and VCSEL based OE-FPGA,” *Proc. SPIE*, vol. 4652C, 2002.
- [6] D. Agarwal and D. A. B. Miller, “Latency in short pulse based optical interconnects,” in *Proc. IEEE LEOS ’01*, 2001, pp. 812–813.
- [7] D. A. B. Miller, “Physical reasons for optical interconnection,” *Int. J. Optoelectron.*, vol. 11, pp. 155–168, 1997.
- [8] G. A. Keeler, B. E. Nelson, D. Agarwal, and D. A. B. Miller, “Skew and jitter removal using short optical pulses for optical interconnection,” *IEEE Photon. Technol. Lett.*, vol. 12, pp. 714–716, 2000.
- [9] C. Debaes, D. Agarwal, A. Bhatnagar, H. Thienpont, and D. A. B. Miller, “High-impedance high-frequency silicon detector response for precise receiverless optical clock injection,” *Proc. SPIE*, vol. 4654, 2002.