IEEE Lasers and Electro-Optics Society 2001 Annual Meeting,
San Diego, California (November 11-15, 2001). Paper ThR3

Latency in Short Pulse based Optical Interconnects

Diwakar Agarwal and David A. B. Miller
Ginzton Laboratory, Stanford University, Stanford, CA 94305

I. Introduction

In connections between and within electronic chips, the total latency (or delay) in sending a logical signal is a very important physical parameter for the performance of the system. If optical interconnects are to relieve the substantial problem of latency in on-chip interconnects, which are projected to be much larger than one clock cycle, the overall latency in the circuits must be on a scale of a few hundred picoseconds or less, [1] so that it is less than the propagation delay in a repeatered electrical connection.

Here we analyze the latency in three optical receiver circuit architectures intended for short optical interconnects. We compare the latency for interconnects using (i) non-return to zero (NRZ) modulation format, and (ii) modulation of short optical pulses. We find that the use of short optical pulses in modulator-based interconnects offers the potential of a significant reduction in latency. In this paper we assume that the GaAs *p-i-n* modulators and diodes are hybrid integrated on 0.25 μm CMOS circuits. The capacitance of these integrated diodes is assumed to be 40 fF and the responsivity 0.5 A/W. Supply voltage in 0.25 μm CMOS is 2.5 V. Simulations are done using hspice simulator.

II. Latency of the Link

Latency of an optical link consists of three components: transmitter latency, the time of flight, and the receiver latency. For modulator-based interconnects the transmitter driver is a buffer chain driving the capacitive load of a modulator. A two buffer chain starting with the minimum size buffer is enough to drive the modulator capacitance ($\sim$ 40 fF) for this CMOS technology. The transmitter delay in this case is $\sim$130 ps. Considering a global on-chip interconnect of length 2 cm, the time of flight delay is $\sim$ 70 ps in free space and somewhat longer in a waveguide or through local optics.

Receiver delay is a major potential source of latency for short optical links. We will consider three different receiver architectures to compare latency and optical power trade-offs. These receivers are shown in Fig. 1. All three receivers are optically differential i.e., they use two differentially modulated optical beams and photo-detectors. The first receiver is an integrating receiver (sense receiver). The input photocurrent is integrated at the gate of the transistors when the clock signal (*clk*) is low and is evaluated when the clock signal goes high. The second receiver is a simple

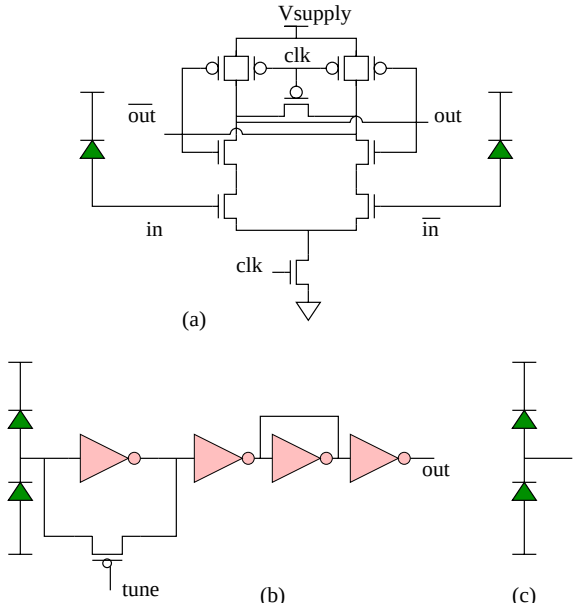

Fig. 1. Schematic representations of receivers considered here: (a) totem-pole diode pair, (b) transimpedance receiver, (c) integrating receiver

transimpedance receiver [2]. The difference photocurrent from two diodes passes through a transimpedance stage that converts input current into voltage and this voltage signal is amplified by the successive stages to a full logic swing. The third receiver is a simple totem-pole diode pair connected to a high impedance node, possibly a buffer. When the data beam is incident on the diodes, the difference in the photocurrent is integrated on the output node and pulls the output to one of the supply rails. Infinite contrast ratio is assumed for both NRZ and short pulse modulation.

A. Integrating Receiver

An integrating receiver consists of a regenerative latch which generates a full swing signal at the output based on the integrated charge at the input. Output of this latch is put into a SR latch to convert from return to zero (RZ) format to NRZ format. The regenerative latch requires a clock input synchronized to the arrival of the data. Delay of the regenerative latch increases exponentially as the input differential charge is reduced. The tradeoff of input optical energy per bit vs. latency for short optical pulse operation is plotted in Fig. 2. Arrival of the pulse was varied with respect to the clock to achieve the shortest possible latency.

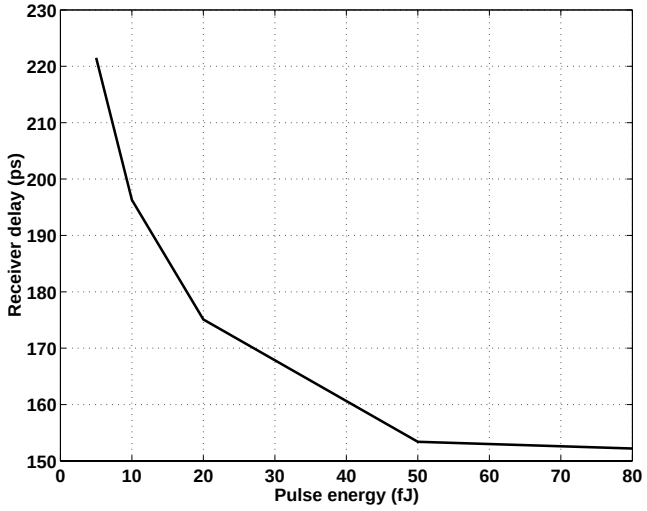

Fig. 2. Power vs. latency using short pulse data input for the integrating receiver

In case of NRZ data input, for maximum efficiency, the data edge should occur at the start of the integrating period. In this case the latency of the receiver is roughly half the bit period. For 1 Gbps operation of the receiver, this latency would be about 500 ps.

B. Transimpedance Receiver

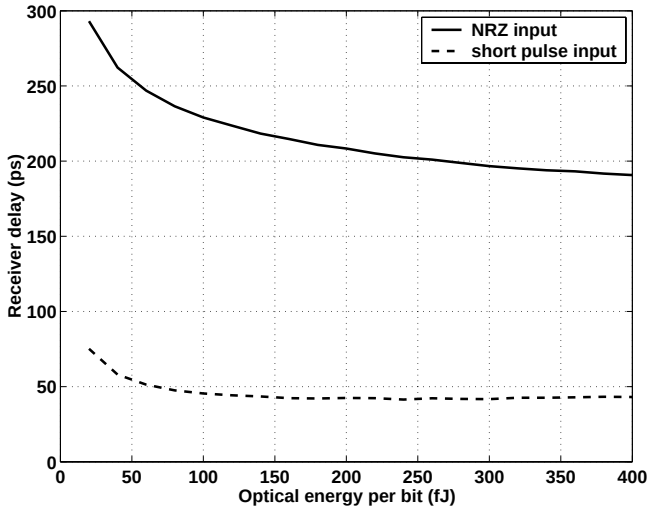

Fig. 3. Power vs. latency for transimpedance receiver for short pulse and NRZ data input

A comparison of the latency for NRZ and short pulse data input for the transimpedance receiver is shown in Fig 3. To calculate the energy per bit for NRZ, 1 Gbps data speed is assumed. The tuning voltage (*tune*) used in the simulation is 0 V. This plot clearly shows that the latency of the receiver is reduced by using short pulse modulation. The short pulses have a sharper rising edge and all the energy is concentrated in a short period. This generates the largest possible swing at the output in the shortest possible time for a given amount of energy. Short pulses, therefore, reduce latency compared to NRZ data.

C. Totem-pole Receiver

In this receiver, large optical power is required to generate full rail-to-rail swings at the output of the diodes, but the latency can be reduced to a minimum possible value by using short pulses. The only latency in this case is due to the carrier transport across the diode, which is nearly velocity saturated because of the high field. We estimate ∼10 ps delay resulting from ~ 10 ps/μm saturated drift velocities for the photodiodes. The optical power required to generate a full swing signal falls linearly as the total capacitance of the front end is reduced. For an output capacitance of 90 fF the minimum optical power required to generate the full swing is 450 fJ. The output capacitance can be reduced by utilizing a silicon on insulator (SOI) process, which reduces the parasitic capacitance to the substrate, and by using metal-semiconductor-metal (MSM) detectors. This receiver would work well at a few select nodes where latency is a primary concern and large optical power can be provided.

For NRZ data input, the output capacitance will be charged in a linear manner. The total energy per bit required to charge the output capacitance to the rail is same as in the short pulse case. In case of short pulses, however, that energy is provided in a very short period giving a corresponding reduction in latency.

III. Conclusions

The use of short pulses reduces the latency of optical interconnect receivers compared to NRZ input due to their sharp rising edges and their concentrated energy in a short period. The totem-pole diode pair receiver with no amplifier provides the lowest possible latency but at the cost of large optical powers. The low latencies possible with this receiver argue for technological development of very low capacitance photodetectors to reduce the optical power required. Latencies achievable in optical receivers may be comparable to the propagation delay time across a chip, making optical interconnect a viable alternative for global on-chip interconnects.

References

[1] Miller D.A.B., "Rationale and Challenges for Optical Interconnects to Electronic Chips", *Proc. IEEE*, v.88, pp. 728 – 749, 2000.

[2] Woodward T.K., Krishnamoorthy A.V., Lentine A.L., and Chirovsky L.M.F., "Optical receivers for optoelectronic VLSI", *IEEE J. on Selected Topics in Quant. Elec.*, v.2, n.1, pp.106-116, April 1996.