Electrical and Optical On-Chip Interconnects in Scaled Microprocessors

Guoqing Chen, Hui Chen, Mikhail Haurylau, Nicholas Nelson, David Albonesi*, Philippe M. Fauchet, and Eby G. Friedman

Department of Electrical and Computer Engineering
University of Rochester, Rochester, New York, 14627

*Department of Electrical and Computer Engineering
Cornell University, Ithaca, New York, 14853

***Abstract*— Interconnect has become a primary bottleneck in integrated circuit design. As CMOS technology is scaled, it will become increasingly difficult for conventional copper interconnect to satisfy the design requirements of delay, power, bandwidth, and noise. On-chip optical interconnect is therefore being considered as a potential substitute for electrical interconnect. Based on predictions of optical device development, electrical and optical interconnects are compared for various design criteria. The critical dimensions beyond which optical interconnect becomes advantageous over electrical interconnect at the 22 nm technology node are approximately one tenth of the chip edge length.**

I. Introduction

In deep submicrometer VLSI technologies, interconnect plays an increasingly important role. Multiple design criteria are considered in interconnect design, such as delay, power, bandwidth, and noise. With technology scaling, it has become increasingly difficult for conventional copper based electrical interconnect to satisfy these requirements. One promising candidate to satisfy these performance objectives is optical interconnect.

The concept of on-chip optical interconnect was first introduced by Goodman in 1984 [1]. Optical interconnects are primarily attractive for global interconnects, such as data buses and clock distribution networks, since electrical/optical and optical/electrical conversion is required. Recently, several comparisons have been made between electrical and optical interconnects [2], [3]. In these papers, the inductive effects of electrical interconnect are ignored, and the parameters characterizing the optical devices are highly approximate.

In this paper, a more comprehensive comparison between optical and electrical interconnects is performed based on a practical prediction of optical device development. The paper is organized as follows. In section II, an RLC model of the delay and power of electrical interconnect is reviewed. The optical data path is introduced in section III. A prediction of the performance characteristics of next generation optical devices is made based on current technology trends. In section IV, electrical and optical interconnects are evaluated based on different criteria. Some conclusions are offered in section V.

II. Electrical Interconnect

Repeater insertion is widely used in submicrometer CMOS technologies to reduce interconnect delay, decrease transition times, and lower crosstalk noise. Numerous papers have been published in this area describing design methodologies that satisfy different design criteria. In this section, an RLC interconnect with repeaters is examined at different technology nodes based on the ITRS [4].

The capacitance and resistance per unit length of the interconnect can be obtained directly from the geometries, where the space between adjacent interconnects is assumed equal to the minimum interconnect width. The interconnect inductance, however, depends upon the distribution of the current return paths which are difficult to estimate before the physical design of the circuit is completed. The sensitivity of a signal waveform to errors in the on-chip inductance, however, is low, and the magnitude of the on-chip inductance is a slowly varying function of the wire geometry [5]. Based on these two characteristics, a fixed value of 0.5 pH/μm [5] is assumed for all of these technology nodes.

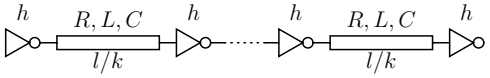

Fig. 1. Repeater insertion in an RLC interconnect.

As shown in Fig. 1, a distributed RLC interconnect with length l is evenly divided into k segments by uniform repeaters. The repeaters are h times as large as a minimum sized repeater, with the output resistance R_{tr0}/h, output capacitance hC_{d0}, and input capacitance hC_{g0}, where R_{tr0}, C_{d0}, and C_{g0} are, respectively, the output resistance, output capacitance, and input capacitance of a minimum sized repeater.

Repeaters are typically implemented as CMOS inverters [6]. In this analysis, the PMOS transistor is assumed to be twice as large as the NMOS transistor. The delay model of the interconnect is an extension of the result from [7] where the repeater output capacitance and input slew effects are considered. The repeater output capacitance is assumed to be the same as the input gate capacitance. The sensitivity of the delay model to this assumption is relatively low. By including the repeater output capacitance, the variable ζ used to characterize inductance effects becomes

$$\zeta = \frac{Rl}{2k}\sqrt{\frac{C}{L}} \cdot \frac{R_T C_T (1 + \frac{C_{d0}}{C_{g0}}) + C_T + R_T + 0.5}{\sqrt{1 + C_T}}, \qquad (1)$$

where $R_T = kR_{tr0}/(hRl)$ and $C_T = hkC_{g0}/(Cl)$. The delay of a single stage interconnect assuming a step input signal can be obtained by curve fitting,

$$t_{ds} = \frac{e^{-2.3\zeta^{1.5}} + 1.48\zeta}{w_n}, \qquad (2)$$

where $w_n = k/\sqrt{Ll(Cl + C_{g0}hk)}$.

With technology scaling, interconnect resistance increases, therefore, ζ is normally greater than 0.5. In this range, the signal transition time at the far end of an interconnect exhibits a linear dependence on ζ,

$$t_{rs} = \frac{4.4\zeta - 1.8}{0.8 w_n}. \qquad (3)$$

The effects of the input transition time on the delay and the far end transition time are treated similarly as in an RC interconnect [8].

The interconnect power models used in this analysis are the same as those models described in [8]. Three degrees of freedom are explored in the design of the electrical interconnect: the wire width, and the number and size of the repeaters. Various combinations are examined to determine the optimal design with respect to a specific criterion.

This research was supported by the National Science Foundation under Contract No. CCR-0304574.

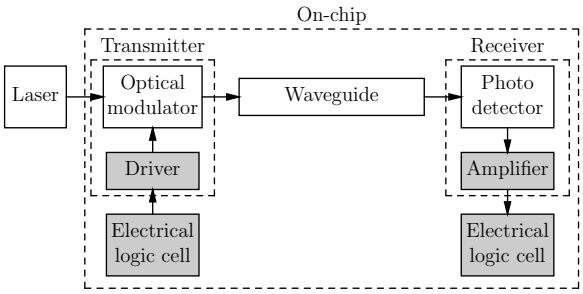

Fig. 2. An on-chip optical interconnect data path.

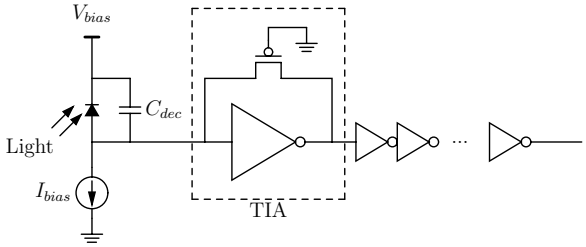

Fig. 3. Circuit model of an optical receiver.

III. On-Chip Optical Data Link

Introducing optical interconnects into VLSI architectures invariablely implies compatibility with CMOS technology. This requirement significantly limits the choice of materials and processes available for fabricating optical components. One of the most significant issue in optical interconnect is the absence of an efficient silicon-based laser that can be monolithically integrated. Only configurations that utilize an external laser are considered. A diagram of an optical interconnect system is shown in Fig. 2. The system consists of three primary optical elements: an off-chip laser, optical modulator, and optical detector. In this paper, low-refractive index polymer waveguides are assumed with an effective index of 1.4 [9].

A. Transmitter

The transmitter is composed of an electro-optical modulator and a driver circuit. The design of a fast and cost-efficient silicon-compatible electro-optical modulator is one of the most challenging tasks on the path towards realizing on-chip optical interconnects. In a modulator, the conversion between electrical and optical signals is performed in two steps. First, certain optical properties of the medium, *e.g.*, the refractive index or absorption coefficient, are changed by the electrical signals. Second, the optical signals are modulated, either in amplitude or phase, by varying the optical properties.

Unstrained bulk crystalline silicon, unfortunately, does not exhibit a linear Pockels effect, and the refractive-index changes due to the Kerr effect are very weak [10]. One of the few suitable mechanisms for varying the refractive index in pure silicon is the free carrier plasma dispersion effect [10]. A number of schemes can be used for free carrier plasma generation. Displacement of carriers is faster than injection schemes as no slow carrier recombination processes are involved. The first MOS capacitor electro-optical modulator based on the carrier displacement effect was demonstrated by Liu *et al.* [11] to operate at frequencies greater than a gigahertz. By design optimization and technology improvements, such as thinning the gate oxide and using an epitaxial over-growth technique, the bandwidth of the modulator is expected to increase to 30 GHz to 40 GHz and the delay will be reduced to less than 20 ps by the year 2016.

Because the device structure used in [11] is a Mach-Zehnder interferometer, the modulator has a large footprint (10 mm long), which also results in an excessive capacitance and hence, increased delay and power consumption of the driver circuits. Simulations and early experiments performed by Barrios *et al.* [10] show that an alternative modulator topology is possible — an optical micro-cavity can drastically decrease the modulator size down to about 10 μm to 30 μm while maintaining the same operating principle and speed. Based on these considerations, the capacitance of the modulator structure is expected to drop below 10 pF within the next few years.

A series of tapered inverters [12] are used to drive the modulator. If the inverter output capacitance is equal to the input gate capacitance, the optimal size ratio between two neighboring inverters is 3.6 [13]. A minimum sized inverter is used as the first stage. The number of stages is $N = \log \frac{C_M}{C_{g0}} / \log 3.6$, where C_M is the modulator capacitance. A delay model of each stage can be obtained from [12], [14].

B. Receiver

The receiver has two components: a photo-detector that converts light into electricity followed by a receiver circuit that amplifies the analog electrical signal to a digital voltage level. A simplified equivalent circuit model is shown in Fig. 3. Because the optical modulator and detector in each optical link operate at the same wavelength, there is a conflict in the requirements of the optical material. In contrast to a modulator, which requires negligible optical absorption, the principle of detector operation relies on the absorption of light. Considering compatibility with CMOS technology, a practical solution is to use a 1.5 μm wavelength light with a SiGe or Ge photo-detector.

In this paper, interdigitated SiGe p-i-n or Metal-Semiconductor-Metal (MSM) detectors are considered due to the fast response and reasonable quantum efficiency of these structures. The signal rise time (response time) of the detector can be expressed as $T_r = \sqrt{T_{tr}^2 + T_{RC}^2}$, where T_{tr} is the time required for the photo-generated carriers to drift to the electrical contact, and T_{RC} is the RC response time of the detector [15]. The 3 dB bandwidth of a detector is $\Delta f_{dec} = 0.35/T_r$. Based on a one pole approximation, the delay of the photo-detector is related to the rise time as $\tau_{dec} = 0.315T_r$. In 2002, an interdigitated Ge p-i-n detector fabricated on a Si substrate with a 3 dB bandwidth of 3.8 GHz at a 1.3 μm wavelength was demonstrated [16]. Several other papers have been published on SiGe detectors and these detectors exhibit similar performance levels, such as [17]. The bandwidth and delay of most of these detectors are limited by the carrier transit time, which can be improved through device optimization. Based on a model proposed by Averine *et al.* [15], the trend in the performance of future detectors is projected. The response time is expected in the near future to drop significantly from tens of picoseconds to a few picoseconds. The cause of this decrease is that present detectors are generally bulky, and a longer time is required for carriers to transit. Effort has therefore been placed on making smaller detectors. Once efficient coupling between the waveguides and the detectors is realized, the size of the detectors is expected to significantly decrease, greatly reducing the response time. This trend, however, is expected to slow and eventually saturate due to fundamental limitations in material properties [18].

The photo-current I_{ph} from the photo-detector is pre-amplified by a trans-impedance amplifier (TIA). The TIA consists of an inverter and a feedback resistor, which is implemented as a PMOS transistor. Additional stages are used to amplify the signal to a digital voltage

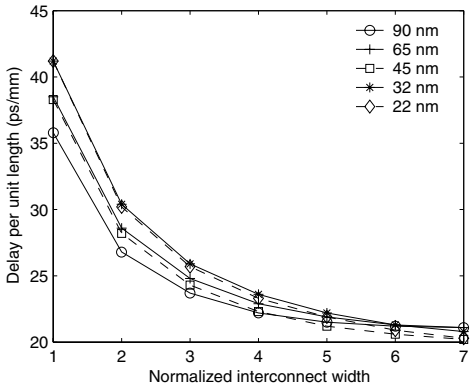

Fig. 4. Minimum delay per unit length as a function of interconnect width.

level. Minimum sized inverters are used as amplifiers to lower the power consumption. A current source I_{bias} is used to bias the input DC current to zero. All of the inverters are assumed to be biased at $V_{dd}/2$. The size of the inverter and the feedback transistor is determined by the bandwidth and noise constraints [19]. The bandwidth requirement of the receiver is assumed to be 0.7 times the bit rate, and the bit error rate (BER) is assumed to be 10^{-15} [19]. For the receiver circuits, the static power dominates and is

$$P_{rec} = W_{TIA} I_{d0} V_{dd} + (I_{bias} V_{dd} + I_{ph} V_{bias})/2 + N_{inv} I_{d0} V_{dd}, \quad (4)$$

where I_{d0} is the saturation drain current of a minimum sized inverter biased at $V_{dd}/2$. W_{TIA} is the size of the TIA normalized to a minimum sized inverter. N_{inv} is the number of additional inverter stages determined by the output swing requirement. The delay of the receiver circuits is obtained by approximating the circuits as a one pole system, $\tau_{cct} = 0.7/(2\pi\Delta f_{req})$, where Δf_{req} is the bandwidth requirement.

IV. Comparison between electrical and optical interconnects

In this section, electrical interconnects are compared with optical interconnects for different design criteria.

A. Delay

The optimal number and size of repeaters along an RLC interconnect can be determined to achieve the minimum delay [7]. This minimum delay can be further decreased by increasing the wire width [20]. The achievable minimum delay per unit length for different wire widths is illustrated in Fig. 4. The interconnect widths are normalized to the minimum wire width W_{min} as predicted by the ITRS. As shown in Fig. 4, scaling has only a small effect on the delay of interconnects with repeaters, consistent with the conclusions from [21]. The decrease in the delay with increasing wire width slows when the wire width exceeds $3W_{min}$. The minimum achievable delay per unit length is approximately a constant – 20 ps/mm for all technology nodes of interest. The delay distribution of a 1 cm optical data path is listed in Table I. The delay of the transmitter is much larger than that of the receiver. By the year 2007 (the 65 nm technology node), optical interconnect is expected to operate faster than electrical interconnect.

B. Power

For electrical interconnects, the power has to be evaluated under specific design requirements, such as delay and bandwidth. A minimum sized wire without repeaters consumes the minimum power; however, this configuration is not practical for global interconnect due to the significant delay and small bandwidth. The power-delay product (PDP) is often used as an effective design criterion [20]. For each wire size, a local minimum PDP can be obtained by adjusting the repeater size and number. From simulations, the global minimum PDP can be achieved with a wire size of $4W_{min}$ to $5W_{min}$ for different technology nodes.

TABLE I

Delay (ps) distribution in a 1 cm optical data path as compared with the electrical interconnect delay.

Year	2004	2007	2010	2013	2016
Technology node	90 nm	65 nm	45 nm	32 nm	22 nm
Modulator driver	83.7	45.8	25.8	16.3	9.5
Modulator	114.0	52.1	30.4	20.0	14.3
Waveguide	46.7	46.7	46.7	46.7	46.7
Photo-detector	1.4	0.5	0.3	0.3	0.2
Receiver amplifier	37.5	16.9	10.4	6.9	4.0
Total optical	283.3	162.0	113.6	90.2	74.7
Electrical	200.0	200.0	200.0	200.0	200.0

TABLE II

Power consumption (mW) in an optical data path.

Year	2004	2007	2010	2013	2016
Technology node	90 nm	65 nm	45 nm	32 nm	22 nm
Transmitter	177.5	18.4	8.6	6.0	5.0
Receiver	0.4	0.3	0.2	0.3	0.3
Total	177.9	18.7	8.8	6.3	5.3

The power consumed by the optical interconnect is almost independent of the interconnect length, since the length is sufficiently short that the optical power loss in the waveguide is negligible. In this paper, only electrical power is evaluated for the optical data path, as listed in Table II. The power consumed by the transmitter dominates the power of the receiver, which is in contrast to the assumption made in [2]. The reason for this difference is that the modulator assumed in this analysis is CMOS compatible. The size as well as the capacitance of the modulator is large and a large driver circuit is therefore needed. Note that there is a significant power decrease from the 90 nm technology node to the 65 nm technology node, which reflects the expected improvements in modulator structures from Mach-Zehnder to a micro-cavity. The PDP of the electrical interconnect is compared with optical interconnect in Fig. 5 for an interconnect length of 1 cm. Note the crossover point between the 65 nm technology node and the 45 nm technology node.

C. Bandwidth density

The maximum bit rate for a single interconnect is assumed to be the clock rate (one bit is transmitted per clock period). With proper design, this bandwidth can be achieved in both the electrical and optical interconnect. The bandwidth density, therefore, is only determined by the interconnect pitch. The waveguide size should be larger than the optical mode size. Based on this limitation, the waveguide pitch is assumed to be 4 μm, much larger than the electrical interconnect pitch, causing a smaller bandwidth density. This drawback, however, can be compensated for by introducing wavelength division multiplexing (WDM). The bandwidth density of different interconnects is compared in Fig. 6. The channel number in a waveguide is assumed to be one in the 90 nm technology node, and to increase by four for each new technology node.

The critical length beyond which optical interconnect overcomes electrical interconnect is plotted in Fig. 7 for different design criteria.

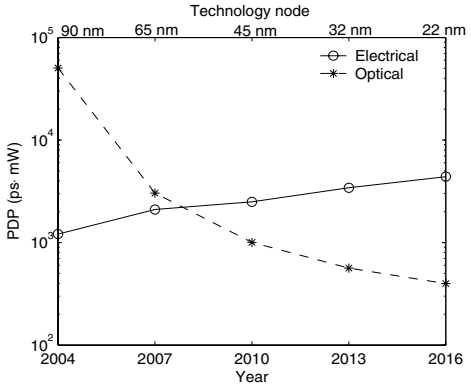

Fig. 5. Comparison of the PDP of electrical and optical interconnects with a length of 1 cm.

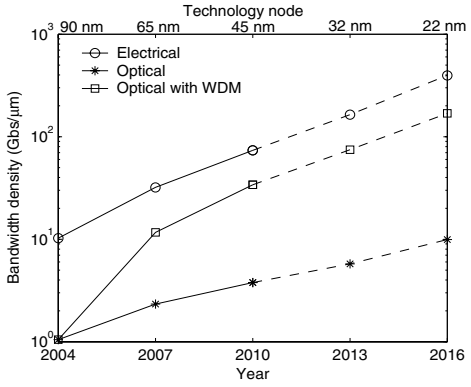

Fig. 6. Comparison of bandwidth density of electrical and optical interconnects.

The lengths are normalized to the edge of the chip die dimension. As shown in Fig. 7, the critical length is approximately one tenth of the chip edge length at the 22 nm technology node.

Note that a fixed optical interconnect design is used in this paper. The optical interconnect can be improved with respect to a specific criterion by further optimizing the circuit. An additional advantage of optical interconnect is the smaller crosstalk noise as compared with electrical interconnect.

V. CONCLUSIONS

Based on a prediction of the performance characteristics of future CMOS compatible optical devices, a comprehensive comparison between electrical and optical on-chip interconnect is presented for different technology nodes. Critical lengths beyond which optical interconnect becomes advantageous are developed for the delay, PDP, and bandwidth density/delay. These lengths are well below the chip die size dimension with technology scaling.

REFERENCES

[1] J. W. Goodman *et al.*, "Optical Interconnects for VLSI Systems," *Proceedings of the IEEE*, Vol. 72, No. 7, pp. 850-866, July 1984.

[2] P. Kapur and K. C. Saraswat, "Comparisons Between Electrical and Optical Interconnects for On-Chip Signaling," *Proceedings of the IEEE International International Interconnect Technology Conference*, pp. 89-91, June 2002.

[3] M. J. Kobrinsky *et al.*, "On-Chip Optical Interconnects," *Intel Technology Journal*, Vol. 8, No. 2, pp. 129-141, May 2004.

[4] *The International Technology Roadmap for Semiconductors*. CA: Semiconductor Industry Association, 2003.

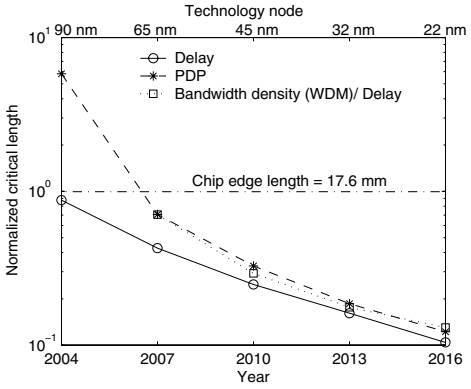

Fig. 7. Normalized critical length beyond which optical interconnect is advantageous over electrical interconnect.

[5] Y. I. Ismail and E. G. Friedman, "Sensitivity of Interconnect Delay to On-Chip Inductance," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 403-406, May 2000.

[6] V. Adler and E. G. Friedman, "Repeater Design to Reduce Delay and Power in Resistive Interconnect," *IEEE Transactions on Circuits and System II: Analog and Digital Signal Processing*, Vol. 45, No. 5, pp. 607-616, May 1998.

[7] Y. I. Ismail and E. G. Friedman, "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 8, No. 2, pp. 195-206, April 2000.

[8] G. Chen and E. G. Friedman, "Low Power Repeaters Driving RC Interconnect with Delay and Bandwidth Constraints," *Proceedings of the IEEE International SOC Conference*, pp. 335-339, September 2004.

[9] L. Eldada and L. W. Shacklette, "Advances in Polymer Integrated Optics," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 6, No. 1, pp. 54-68, January/February 2000.

[10] C. A. Barrios, V. R. Almeida, and M. Lipson, "Low-Power-Consumption Short-length and High-Modulation-Depth Silicon Electrooptic Modulator," *Journal of Lightwave Technology*, Vol. 21, No. 4, pp. 1089-1098, April 2003.

[11] A. Liu *et al.*, "A High-Speed Silicon Optical Modulator Based on a Metal-Oxide-Semiconductor Capacitor," *Nature*, Vol. 427, pp. 615-618, February 2004.

[12] B. S. Cherkauer and E. G. Friedman, "A Unified Design Methodology for CMOS Tapered Buffers," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 3, No. 1, pp. 99-111, March 1995.

[13] N. Hedenstierna and K. O. Jeppson, "CMOS Circuit Speed and Buffer Optimization," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 6, No. 2, pp. 270-281, March 1987.

[14] T. Sakurai and A. R. Newton, "Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas," *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 2, pp. 584-594, April 1990.

[15] S. V. Averine, Y. C. Chan, and Y. Lam, "Geometry Optimization of Interdigitated Schottky-Barrier Metal-Semiconductor-Metal Photodiode Structures," *Solid-State Electronics*, Vol. 45, No. 3, pp. 441-446, March 2001.

[16] J. Oh *et al.*, "Interdigitated Ge p-i-n Photodetectors Fabricated on a Si Substrate Using Graded SiGe Buffer Layers," *IEEE Journal of Quantum Electronics*, Vol. 38, No. 9, pp. 1238-1241, September 2002.

[17] J. Oh, S. K. Banerjee, and J. Campbell, "Metal-Germanium-Metal Photodetectors on Heteroepitaxial Ge-on-Si with Amorphous Ge Schottky Barrier Enhancement Layers," *IEEE Photonics Technology Letters*, Vol. 15, No. 5, pp. 745-747, May 2003.

[18] L. Pavesi and D. J. Lockwood, *Silicon Photonics*. NY: Springer, 2004.

[19] P. Kapur, "Scaling Induced Performance Challenges/Limitations of On-Chip Metal Interconnects and Comparisons with Optical Interconnects," Ph.D. Dissertation, Stanford University, Stanford, CA, 2002.

[20] M. A. El-Moursy and E. G. Friedman, "Optimum Wire Sizing of RLC Interconnect with Repeaters," *Proceedings of the IEEE Great Lakes Symposium on VLSI*, pp. 27-32, April 2003.

[21] R. Ho, K. W. Mai, and M. A. Horowitz, "The Future of Wires," *Proceedings of the IEEE*, Vol. 89, No. 4, pp. 490-504, April 2001.