On-Chip Optical Interconnect for Reduced Delay Uncertainty

Guoqing Chen, Hui Chen*, Mikhail Haurylau, Nicholas A. Nelson,
David H. Albonesi**, Philippe M. Fauchet, and Eby G. Friedman
Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York, 14627
*Department of Physics and Astronomy, University of Rochester, Rochester, New York, 14627
**School of Electrical and Computer Engineering, Cornell University, Ithaca, New York, 14853

***Abstract*— Interconnect has become a primary bottleneck in the integrated circuit design process. As CMOS technology is scaled, the design requirements of delay, power, bandwidth, and noise due to the on-chip interconnects have become increasingly stringent. New design challenges are continuously emerging, such as delay uncertainty induced by process and environmental variations. It has become increasingly difficult for conventional copper interconnect to satisfy a variety of design requirements. On-chip optical interconnect has been considered as a potential partial substitute for electrical interconnect. In this paper, predictions of the performance of CMOS compatible optical devices are made based on current state-of-the-art optical technologies. Based on these predictions, the delay uncertainty in electrical and optical interconnects is analyzed, and shown to affect both the latency and bandwidth of the interconnect. The two interconnects are also compared for latency, power, and bandwidth density.**

I. Introduction

In deep submicrometer VLSI technologies, it has become increasingly difficult for conventional copper based electrical interconnect to satisfy the design requirements of delay, power, and bandwidth. One promising candidate to solve this problem is optical interconnect. Based on a practical prediction of optical device development, a comprehensive comparison between optical and electrical interconnects is described for different technology nodes. Delay uncertainty, as an emerging critical design criterion, particularly in digital systems, is shown to affect both the interconnect latency and bandwidth. The rest of this paper is organized as follows. In section II, models of electrical and optical interconnects are introduced. In section III, delay uncertainty in these interconnects is analyzed. In section IV, electrical and optical interconnects are compared for latency, power, and bandwidth density. Some conclusions are offered in section V.

II. Electrical and optical interconnects

In the analysis of electrical interconnect, a distributed RLC interconnect with repeaters is considered, as shown in Fig. 1. Three degrees of freedom are explored: the wire width, and the number and size of the repeaters. Various structures are examined to determine the optimal interconnect to achieve the minimum delay. The minimum achievable delay per unit length is approximately in the range of 20 ps/mm to 22 ps/mm for all technology nodes of interest. The optimal wire width is chosen as $7W_{min}$ for each technology node [1], where W_{min} is the minimum wire width predicted in the ITRS [2].

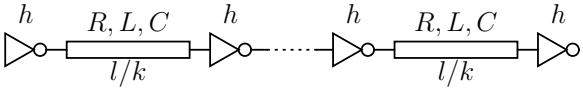

Fig. 1. Repeater insertion in an RLC interconnect.

Introducing optical interconnects into VLSI architectures requires compatibility with CMOS technology. Due to the absence of an efficient silicon-based laser, only those configurations that utilize an external laser as a light source are considered. A diagram of an optical interconnect system is shown in Fig. 2. A transmitter is used to convert an electrical signal to a light signal, which is composed of a modulator and a driver circuit. The development of a fast and cost efficient CMOS compatible electro-optical modulator is one of the most challenging tasks on the path towards realizing on-chip optical interconnects. In this paper, a predictive modulator model [3] is used which is based on an MOS capacitor micro-resonator structure. Polymer or silicon-on-insulator (SOI) waveguides are used to transmit the optical signal. The optical signal is converted to an electrical signal at the receiver. The receiver has two components: a SiGe metal semiconductor metal (MSM) detector and an amplifier. Models of different components within the optical data path are provided in [1]. Unlike electrical devices, optical devices are not readily scalable due to the light wavelength constraint. The performance and integration ability of optical devices, however, are expected to be further improved by technology innovations and structural optimization.

III. Delay uncertainty

Delay uncertainty is caused by process and environmental variations. Variations from the environment include power/ground noise, temperature fluctuations, and crosstalk coupling. In this paper, all of the variations are assumed to be random with a normal distribution and independent unless explicitly indicated.

Process variations include both die-to-die and within-die variations. Temperature variations also exhibit a similar behavior. The average on-chip temperature is different from die

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

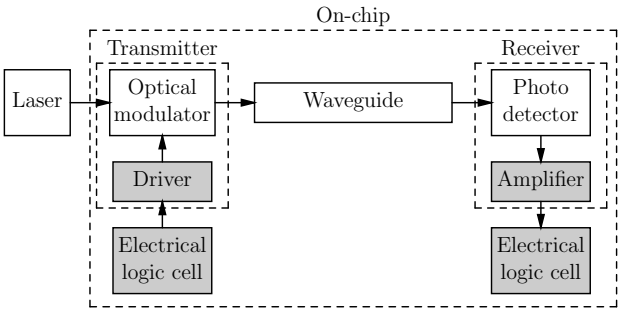

Fig. 2. An on-chip optical interconnect data path.

to die due to the ambient environment and local circuit activity. Since different on-chip blocks typically dissipate different amounts of power, the temperature is also non-uniform within a die [4]. The within-die variations at different locations are generally correlated according to the physical separation. Since the focus of this paper is on comparing electrical and optical global interconnects which cross different regions of an IC, within-die spatial correlation effects need to be considered. The spatial correlation coefficient is modeled as [5]

$$\rho_{cor}(x) = \begin{cases} 1 - \frac{x}{X_L}(1-\rho_B) & x \le X_L, \\ \rho_B & x > X_L, \end{cases} \quad (1)$$

where x is the physical separation between the variations, ρ_B is the correlation coefficient of the die-to-die variations, and X_L is related to the gradient of the systematic within-die variations. In this paper, ρ_B is assumed to be 0.5 for process and temperature variations, *i.e.* the variations are caused equally by die-to-die and within-die variations. X_L is assumed to be 10 mm for MOS device variations, temperature variations, and power/ground noise, 2 mm for the interconnect height and inter-layer dielectric thickness, and 5 mm for the interconnect width and spacing [6].

The process variations considered in the MOS transistors are the effective channel length, gate oxide thickness, channel doping concentration, and drain/source resistance. By using analytic models of threshold voltage [7], [8], carrier mobility [9], and saturation drain current [10], the variation of the transistor performance can be analyzed.

For electrical interconnects, process variations occur in the geometric parameters, such as the wire width, wire height, space between wires on the same level, and thickness of the inter-layer dielectric. The variation in resistance R is primarily determined by geometric parameters and the temperature of the interconnect,

$$R = \frac{\rho_0 l}{WH}[1 + \beta(T - T_0)], \quad (2)$$

where β is the temperature coefficient of resistivity, T_0 is the reference temperature, and ρ_0 is the resistivity at T_0. The interconnect capacitance can be expressed as $C_{gnd} + 2\eta C_c$, where C_{gnd} is the ground capacitance and C_c is the coupling capacitance to the neighboring wires. The switching factor η models the Miller effect due to switching activities on neighboring wires [11]. It is shown in [12] that the effective wire inductance can be described as $L = L_{self} + \sum_i \xi_i M_i$, where L_{self} is the partial self inductance and M_i is the partial mutual inductance between wire i and the wire of interest. ξ_i is a coefficient which depends upon the signal switching patterns and wire capacitances. To simplify the problem, the inductance is modeled in this paper as $L = L_0(1 + \xi)$, where L_0 is a typical value of inductance of 0.5 pH/μm [13] and ξ is used to model the effect of coupling on the effective inductance.

Those parameters considered to vary and the corresponding 3σ values can be found in [1]. The absolute variations in the optical devices are assumed to be the same as in a similar electrical structure, since the same technology is applied to fabricate these optical devices. Among these parameters, three kinds of correlations are considered. First, power and ground are inversely correlated with a correlation coefficient of -0.5. Second, assuming a fixed wire pitch, the interconnect width and space are inversely correlated with a coefficient of -1. Third, ξ is assumed to be correlated to η with a correlation coefficient of -0.3. This assumption is based on the observation that oppositely switching neighbors result in a greater effective capacitance (*i.e.*, a greater η), and a smaller effective inductance (*i.e.*, a smaller ξ), since neighboring wires provide nearby current return paths. Unlike the effective capacitance, which is only related to the immediate neighbors, the effective inductance depends on the neighboring wires over a long distance, making the correlation between η and ξ fairly weak.

The electrical interconnect is divided into a number of segments by repeaters. In each segment, process, temperature, and power/ground variations are assumed to be uniformly distributed, *i.e.*, ρ_{cor} inside a segment has a value of one. ρ_{cor} between different segments is determined by (1). η and ξ, however, are assumed to be uniform along the total length of the interconnect, since in a bus structure wires often experience the same neighboring coupling environment over the total length of the line.

For an optical interconnect system, although the waveguide crosses a long distance, geometric variations are assumed to be uniform across the total length. This assumption overestimates the delay uncertainty, since those independent components of variations in different parts of the waveguide can average out, producing a smaller delay uncertainty. This overestimation, however, does not affect the primary conclusions of this paper, since delay uncertainty caused by the waveguide is small as compared with other parts of the system. Although parameter variations in different parts of an optical interconnect may be correlated, the effects of these variations on delay uncertainty are different due to different operational mechanisms. In this paper, the delay uncertainty generated at different parts of the optical data path is assumed to be independent, resulting in the following expression for the standard deviation of the total delay,

$$\sigma_{optical} = \sqrt{\sigma_{drv}^2 + \sigma_{mod}^2 + \sigma_{wav}^2 + \sigma_{dec}^2 + \sigma_{amp}^2}\,. \quad (3)$$

Based on these assumptions, the delay uncertainty of both electrical and optical interconnect is analyzed. The delay and

TABLE I

DELAY AND 3σ VALUE OF A 1 cm OPTICAL DATA PATH.

Tech. node	90 nm		65 nm		45 nm		32 nm		22 nm	
	Delay (ps)	3σ (%)	Delay (ps)	3σ (%)	Delay (ps)	3σ (%)	Delay (ps)	3σ (%)	Delay (ps)	3σ (%)
Mod. driver	37.3	20.9	26.5	20.4	16.6	23.5	10.3	29.1	5.2	40.4
Modulator	40.0	67.0	40.0	51.0	40.0	41.0	40.0	32.0	40.0	27.0
Waveguide	49.3	1.1	49.3	0.8	49.3	0.5	49.3	0.2	49.3	0.1
Detector	2.5	5.6	1.1	21.9	0.6	14.1	0.5	9.3	0.4	7.1
Amplifier	34.0	10.6	13.5	23.8	8.7	17.6	5.7	15.8	3.4	15.0
Total optical	163.1	17.3	130.4	16.4	115.2	14.7	105.8	12.5	98.3	11.2

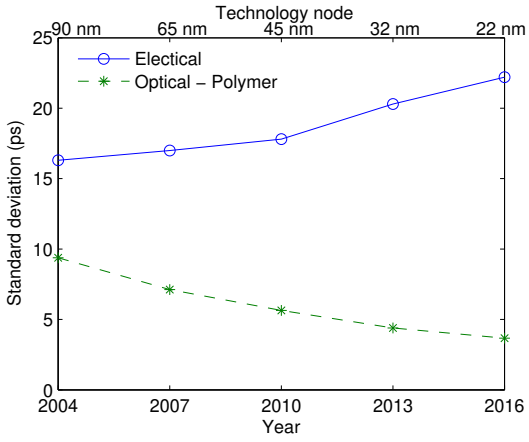

Fig. 3. Comparison of standard deviation of electrical and optical interconnect delays.

3σ value for different portions of a 1 cm optical data path are listed in Table I. A comparison of the standard deviation of the delay of electrical and optical interconnect is shown in Fig. 3. The delay uncertainty of optical interconnect is expected to be lower in future technology nodes. The delay uncertainty of electrical interconnect, in contrast, is expected to slowly increase in future technology nodes due to the larger number of inserted repeaters.

IV. DELAY, POWER, AND BANDWIDTH DENSITY

In order for the data to be correctly latched at the receiving register, specific setup and hold constraints should be satisfied, as shown in Fig. 4. In this paper, the timing budget assigned to T_{setup} and T_{hold} is assumed to be 20% of the clock period, *i.e.*, the delay uncertainty T_{un} cannot exceed 80% of the clock period. If this requirement is not satisfied, pipeline registers are inserted such that the timing requirements of each stage are satisfied. The actual delay of the interconnect considering delay uncertainty is

$$T_{total} = m(T_{max} + T_{setup} + T_{C-Q}), \tag{4}$$

where m is the number of register stages, T_{C-Q} is the clock-to-data delay, and T_{max} is the maximum delay of each stage. $T_{setup}+T_{C-Q}$ is also assumed to be 20% of the clock period. Since register-like devices cannot be inserted into an optical data path, the delay uncertainty provides an upper bound on the optical channel bandwidth, $B_{optical} = \frac{1}{T_{bit}} \leq \frac{0.8}{T_{un}}$.

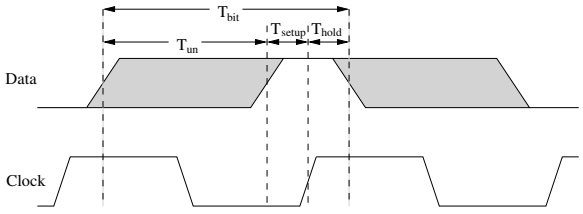

Fig. 4. Timing diagram of data and clock waveforms.

TABLE II

LATENCY COMPARISON BETWEEN ELECTRICAL AND OPTICAL INTERCONNECTS.

Year		2004	2007	2010	2013	2016
Technology node		90 nm	65 nm	45 nm	32 nm	22 nm
Electrical	Delay (ps)	311.9	313.2	291.3	312.0	317.8
	# of register stages	1	2	2	4	7
	# of clock cycles	2	3	5	7	12
Optical Polymer	Delay (ps)	238.9	173.3	145.4	127.7	114.8
	# of register stages	1	1	1	1	1
	# of clock cycles	1	2	2	3	4
Optical SOI	Delay (ps)	291.6	226.0	198.1	180.4	167.5
	# of register stages	1	1	1	1	1
	# of clock cycles	2	2	3	4	6

By considering the effect of the registers, the delay of the electrical interconnect and optical interconnect with a polymer or SOI waveguide is listed in Table II. As listed in this table, the actual delay of the electrical interconnect remains approximately fixed for all of those technology nodes. The delay of the optical interconnect, however, decreases with future technology nodes due to the higher performance of the electrical circuits in the modulator driver and receiver amplifier. The polymer waveguide is advantageous as compared to an SOI waveguide in terms of latency due to the higher light speed in the polymer.

The electrical interconnect power models used in this analysis are the same as those models used in [14]. The power dissipated by the registers can be estimated by scaling a typical master-slave D flip-flop. The power due to the registers is negligible as compared to the power of the interconnects. For optical interconnect, only the electrical portion of the power is evaluated, which is almost independent of the interconnect length, since the length is sufficiently short such that the optical power loss in the waveguide is negligible. The power dissipated by the electrical and optical interconnect systems is compared in Table III. Both the electrical and optical interconnect power increases due to higher clock frequencies and greater leakage current. In optical interconnect, the power consumed by the transmitter dominates the power of the receiver, which is in contrast to the assumption made in [15]. The reason for this difference is that the modulator assumed in this analysis is CMOS compatible, with a large capacitance, requiring a large driver.

Bandwidth density is an effective criterion for evaluating the ability to transmit data through a unit width. The maximum bit rate for a single interconnect is assumed to be the clock rate. From [1], the optimal interconnect width is $7W_{min}$,

TABLE III

POWER (mW) OF OPTICAL AND ELECTRICAL INTERCONNECTS.

Technology node	90 nm	65 nm	45 nm	32 nm	22 nm
Transmitter	0.9	1.9	3.4	5.9	11.2
Receiver	0.5	0.5	0.3	0.3	0.3
Total optical	1.4	2.4	3.7	6.2	11.5
Electrical	9.8	16.9	21.7	33.4	45.3

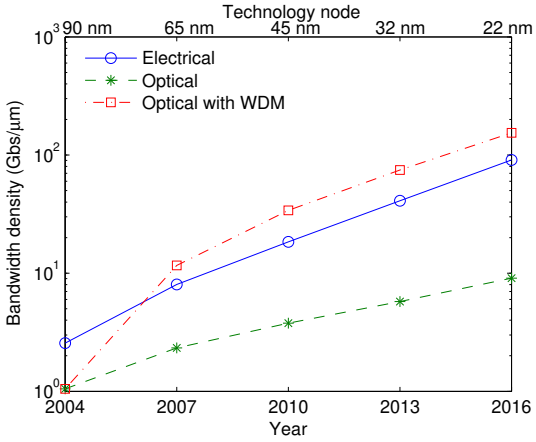

Fig. 5. Bandwidth density of electrical and optical interconnects.

corresponding to a pitch of $8W_{min}$. Requiring the waveguide size to be larger than the optical mode size, the waveguide pitch is assumed to be 4 μm. Single wavelength optical interconnects are not beneficial if high bandwidth density is desired. The bandwidth of optical interconnects, however, can be significantly improved by introducing wavelength division multiplexing (WDM). The bandwidth density of different interconnects is compared in Fig. 5. For optical interconnect with WDM, the channel number in a waveguide is assumed to be one at 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. 6 for different design criteria. The lengths are normalized to the edge of the chip die dimension. As shown in Fig. 6, the critical length is approximately one tenth of the chip edge length at the 22 nm technology node.

V. CONCLUSIONS

A prediction of the performance characteristics of future CMOS compatible optical devices is described in this paper. Based on this prediction, electrical and optical on-chip interconnects are compared for various design criteria at different technology nodes. For an interconnect length of 10 mm, optical interconnect is shown to exhibit a smaller delay uncertainty and this benefit is expected to increase with technology improvements. Critical lengths beyond which optical interconnect becomes advantageous are presented. These lengths are well below expected chip die size dimensions.

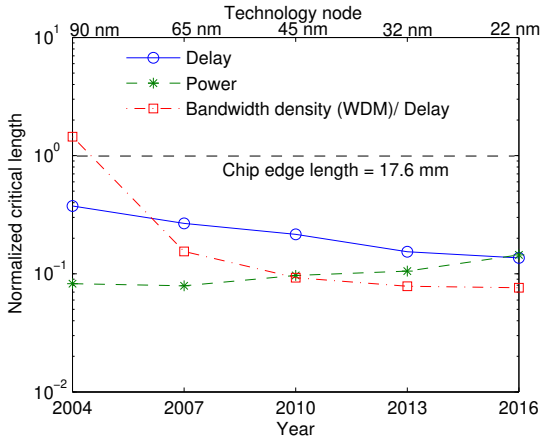

Fig. 6. Normalized critical length beyond which optical interconnect is advantageous over electrical interconnect.

REFERENCES

[1] G. Chen *et al.*, "Predictions of CMOS Compatible On-Chip Optical Interconnect," *Integration, the VLSI Journal*, Vol. 40, No. 4, pp. 434–446, July 2007.

[2] *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2003.

[3] M. Haurylau *et al.*, "Closed-Form Model of a Capacitor-Based Electro-Optical Modulator," *IEEE Transactions on Electron Devices*, 2007 (in review).

[4] A. H. Ajami, K. Banerjee, and M. Pedram, "Modeling and Analysis of Nonuniform Substrate Temperature Effects on Global ULSI Interconnects," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 6, pp. 849–861, June 2005.

[5] P. Friedberg *et al.*, "Modeling Within-Die Spatial Correlation Effects for Process-Design Co-Optimization," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 516–521, March 2005.

[6] Y. Cao *et al.*, "Design Sensitivities to Variability: Extrapolation and Assessments in Nanometer VLSI," *Proceedings of the IEEE ASIC/SOC Conference*, pp. 411–415, September 2002.

[7] Z. Liu *et al.*, "Threshold Voltage Model for Deep-Submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, Vol. 40, No. 1, pp. 86–95, January 1993.

[8] G. F. Niu, G. Ruan, and R. M. M. Chen, "Further Comments on 'Threshold Voltage Model for Deep-Submicrometer MOSFET's' and Its Extension to Subthreshold Operation," *IEEE Transactions on Electron Devices*, Vol. 43, No. 12, pp. 2311–2312, December 1996.

[9] N. Arora, *MOSFET Models for VLSI Circuit Simulation – Theory and Practice*, Springer, NY, 1993.

[10] N. G. Einspruch, *VLSI Electronics Microstructure Science – Advanced MOS Device Physics*, Academic Press, San Diego, California, 1989.

[11] A. B. Kahng, S. Muddu, and E. Sarto, "On Switch Factor Based Analysis of Coupled *RC* Interconnects," *Proceedings of the IEEE/ACM Great Lakes Symposium on VLSI*, pp. 79–84, June 2000.

[12] Y. Cao *et al.*, "Switch-Factor Based Loop *RLC* Modeling for Efficient Timing Analysis," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 13, No. 9, pp. 1072–1078, September 2005.

[13] 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.

[14] G. Chen and E. G. Friedman, "Low Power Repeaters Driving *RC* and *RLC* Interconnects with Delay and Bandwidth Constraints," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 14, No. 2, pp. 161–172, February 2006.

[15] 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.