PERFORMANCE COMPARISON BETWEEN COPPER, CARBON NANOTUBE, AND OPTICS FOR OFF-CHIP AND ON-CHIP INTERCONNECTS

A DISSERTATION
SUBMITTED TO
THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTORAL OF PHILOSOPHY

Hoyeol Cho
March 2007

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Krishna C. Saraswat) Principal Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(James S. Harris)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Pawan Kapur)

Approved for the University Committee on Graduate Studies.

Abstract

For more than 30 years, the performance of silicon integrated circuits has improved at an astonishing rate. The number of functions per chip has grown exponentially, dramatically bringing down the cost per function. However, for the first time the relentless scaling paradigm is threatened by fundamental limits including excessive power dissipation, insufficient communication bandwidth, and signal latency. Many of these obstacles stem from the physical limitation of Cu-based electrical wires, making it imperative to examine alternate interconnect schemes for future ICs. The two most important novel potential candidates are optical and carbon nanotube (CNT)-based interconnects.

Optical interconnect due to its high bandwidth, low signal attenuation and cross talk, is an ideal candidate to tackle the challenges imposed by electrical wiring for both off-chip and possibly on-chip application. Because modern ICs require an ever-increasing system bandwidth and have a large power density, a realistic study of performance comparison between electrical and optical interconnects is of paramount importance. In addition, such a comparison framework aids in setting clear goals on the requirements of opto-electronic devices to deliver performance superior to that of their electrical counterparts. For on-chip applications, optical interconnects can potentially reduce latency and provide high-bandwidth at relatively low power. However, an optical waveguide, has a relatively larger size (pitch~0.6μm), making it difficult to provide high bandwidth density. This can be mitigated using wavelength division multiplexing (WDM). CNT interconnects, on the other hand, have the flexibility of being implemented in the same size scale as the existing Cu on-chip wires, hence possibly can provide high bandwidth density. In addition they have the advantage of having a large electron mean free path, hence low resistance. This can result in low latency compared to its Cu/low-K counterpart.

In this dissertation, for off-chip applications, we compare high speed optical and electrical interconnects using power vs. bandwidth. We find that beyond a critical length, power optimized optical interconnects dissipates less power compared to high-speed electrical signaling schemes. Beyond the 32nm technology node with its commensurate bandwidth, optical interconnect becomes favorable for the distances less than 10cm for inter-chip communications. This analysis of the impacts of device/system parameters on critical length gives system/device designers the evaluation framework of the performance of the system. We also examine two competing transmitters, the vertical cavity surface emitting laser (VCSEL) and the quantum well modulator (QWM), for optical links. We identify lower capacitance (<15fF) of both modulators and the photo-detectors as the critical parameter to render QWM more power efficient. Furthermore, we present an optimization methodology of minimizing total optical link power, and obtain the optimized modulator design parameters. As bit rate increases, the optimum voltage swing exceeds the stipulated supply voltage at the 65nm technology node. This results in about 46% power penalty at 25Gb/s.

Finally, for on-chip applications, we compare CNT and optical interconnects with Cu interconnects using both commonly used metrics: latency and power and a compound metric which captures system requirements more efficiently. The necessity of compound metrics is motivated by the fact that a larger bandwidth and a smaller latency can be obtained using more area resources. Hence area normalization is necessary. In addition, the total power budget can also be used to increase the aggregate bandwidth, making power normalization also imperative. Hence, we use bandwidth density/latency/power as our compound metric. In the future, because of multi-core architecture, designers care about bandwidth density, latency, and power dissipation of global communication. We extensively examine the impact of device parameters-modulator and detector capacitances for optics, materials parameters-mean free path and packing density for CNTs, and system parameters-global clock frequency and switching activity on both commonly used and compound metrics. We

find that at the 22nm technology node small detector and modulator capacitances for optical interconnects (~10fF) yield superior, at least comparable, performance with CNTs (electron mean free path of 0.9μm) and Cu for greater than 35% and 20% switching activity, respectively. However, improving mean free path of CNTs (~2.8μm) increases this crossover switching activity to 80%.

Acknowledgments

The title page of this doctoral thesis lists my name as the single author, although that is a deception. The ideas and concepts in the pages that follow arose from the work of many people, and to them I owe my thanks.

I want to express my heartfelt gratitude toward my advisor Prof. Krishna C. Saraswat for providing me invaluable guidance during my stay at Stanford. He provided me tremendous freedom to pursue various ideas of my interest as well as nudged me in the right direction with his vision. I benefited enormously from his wealth of experience in the technical as well as non-technical areas. I also appreciate my associate advisor, Prof. James S. Harris for his assistance and input to my research and dissertation. I was very fortunate to have Dr. Pawan Kapur as my research mentor and one of reading committee members. The valuable discussion with him improves my research ability and makes it fruitful. I also want to thank Prof. Phillip H. S. Wong for agreeing to be on my oral committee and Prof. Christos Kozyrakis for being the chair of my orals.

I've had the pleasure of working with many of Krishna's graduate students. Much of my learning at Stanford was a result of numerous stimulating discussions I had with them. For this and for all the other things which the limited space does not permit me to write about, I would like to thank to Abhijit, Ali, Amar, Crystal, Dong-Hyun, Duygu, Hoon, Hyun-Yong, Jin-Hong, Kunhan, Kyung-Hoae, Rohit, Sham, Salves, Sham, and Tejas. I also would like to put a special mention to Henry for helping building chips.

The staffs in Stanford Nano-Fabrication (SNF) and Ginzton Lab have been helpful, knowledgeable, and supportive, as needed. For this, I am very grateful. Tom

Carver was especially very kind and patient. I also would like to special thanks to Irene Sweeney for her patience and for help with administrative issues.

There are no words to thank my parents. I am what I am and where I am because of their innumerable sacrifice, unconditional love, continuous hard work and unwavering faith in me. I also want to thank all my other family members, especially my brother, Kwanyeol, taught me how to think and how to be curious about how and why things worked. This started me on this direction from my childhood.

Finally, I want to thank one of the most important person in my life, my wife Jeongae, who has always showed unflagging support, optimism, and patience. Without her encouragement, support, sacrifice, and her faith, I would never have finished this work and so humbly I dedicate it to her.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

For more than 30 years, the performance of silicon integrated circuits has improved at an astonishing rate. The number of functions per chip has grown exponentially, dramatically bringing down the cost per function. However, for the first time the relentless scaling paradigm is threatened by fundamental limits including excessive power dissipation, insufficient communication bandwidth, and signal latency. Many of these obstacles stem from the physical limitation of Cu-based electrical wires [1]-[3], making it imperative to examine alternate interconnect schemes for future ICs. The two most important novel potential candidates are optical and carbon nanotube (CNT)-based interconnects.

In the short distance interconnect applications (<100m), the interconnect hierarchy is typically divided into the following categories in the order of progressively smaller length spans: cabinet level (1-100m), backplane level between

boards (10cm-1m), and chip to chip on a board (<10cm) and on-chip (<2cm) communication [4]. Optical interconnect due to its high bandwidth, low signal attenuation and cross talk, is an ideal candidate to tackle the challenges imposed by electrical wiring for both off-chip and possibly on-chip application. Because the modern ICs require an ever-increasing system bandwidth and have a large power density, a realistic study of performance comparison between electrical and optical interconnects is of paramount importance. In addition, such a comparison framework aids in setting clear goals on the requirements of opto-electronic devices to deliver a superior performance than their electrical counterparts. For on-chip application, optical interconnects can potentially reduce latency, provide high-bandwidth at relatively low power. However, an optical waveguide, has a relatively larger size (pitch~0.6μm), making it difficult to provide high bandwidth density. This can be mitigated using wavelength division multiplexing (WDM). CNT interconnects, on the other hand, have the flexibility of being implemented in the same or even smaller size scale as the existing Cu on-chip wires, hence possibly can provide high bandwidth density. In addition they have the advantage of having a large electron mean free path, hence low resistance. This can result in low latency compared to Cu/low-K counterpart.

1.2 Performance Metrics for Interconnects

The proper performance metrics are imperative for comparing between different interconnect technologies. The appropriateness of the metrics depends on whether they are on or off-chip interconnect and on the particular function that the interconnects perform. Bandwidth, power, area (wire pitch), and delay (latency) are commonly used performance metrics. These metrics would henceforth be referred to

as primary metrics. Table 1.1 presents 1) what are the interconnect technologies that we investigate for different interconnect levels, 2) how the primary metrics are related with interconnect levels and the system requirement. For off-chip interconnect, we compare Cu-based electrical and optical interconnect, while for on-chip interconnect, we compare Cu, CNT, and optics.

Table 1.1 Performance metrics for off-chip and on-chip interconnects

<table>
<tr><th></th><th>Off-chip</th><th>On-chip</th></tr>
<tr><td>Interconnect technology</td><td>Cu, Optics</td><td>Cu, CNT, Optics</td></tr>
<tr><td>Bandwidth</td><td>√</td><td>Fixed bandwidth</td></tr>
<tr><td>Power</td><td>√</td><td>√</td></tr>
<tr><td>Area (wire pitch)</td><td rowspan="2">Comparable</td><td>√ (bandwidth density)</td></tr>
<tr><td>Delay (latency)</td><td>√</td></tr>
<tr><td>Proper performance metrics</td><td>Power vs. Bandwidth</td><td>Compound metric: bandwidth density/power/latency</td></tr>
</table>

For off-chip interconnects, the system bandwidth requirement increases drastically as technology scales (ITRS [6]). This inherently increases power dissipation. Further, the chip power increases with more functionality and number of integrated transistors, while chip area does not change appreciably, resulting in a ever-increasing power density. Thus, it is imperative to reduce the power per link for a given bandwidth. The area and latency between Cu-based electrical and optical interconnect are comparable. Off-chip electrical interconnect use the printed circuit board (PCB) trace, which is in the LC regime requiring wire pitch comparable to a waveguide for optical interconnect. In the LC regime, the propagation of signals is

effectively at the velocity of light in the medium as in the case of optical interconnect. Based on this discussion, as well as on the fact that the interconnect throughput and the power dissipation are of the greatest interest in most applications, we compare the various off-chip interconnects in terms of power and bandwidth.

On the other hand, for on-chip interconnects, the system bandwidth is determined by the global clock frequency. Clock frequency has been increasing in the past but it is leveling-off and might saturate at about 10-20Gb/s. For on-chip interconnects, the wire pitch of Cu and CNT, which is in the RC regime (~0.1μm), is smaller than that of optical waveguides (~0.6μm). Thus, a normalization with respect to wire pitch, leading to bandwidth density would be a better performance metric. However, bandwidth density, latency and power by themselves are not necessarily the best metrics for comparing different interconnects, as a larger bandwidth and a small latency can be obtained using more area resources. Naeemi et al. [7] proposed bandwidth density/latency; however, the total power budget is another factor which can be used to increase the aggregate bandwidth, hence a normalization with respect to power is also imperative. In addition, a normalization with latency is important as it can contribute to the communication bottleneck depending on the nature of the data traffic. Thus, bandwidth density per latency per power, severs as a fair compound metric to compare various interconnect technologies. For this metric, bandwidth density per latency is the measure of interconnect performance and power is the price to attain that performance. The reason this metric is important is because for global communication between blocks or cores, the designers care about the bandwidth density, the latency of the links, and how much power they need to expend to obtain these links.

1.3 This Work and Dissertation Organization

Keeping the proper performance metrics described in previous section in mind, we have divided this dissertation into two parts. In the first part, which includes Chapter 2-5, we deal with realistic modeling of off-chip electrical and optical interconnects for their performance comparison, optimizing the transmitter and receiver devices. In addition, we describe the design of a prototype chip, handling chip to chip interconnections. In the second part, Chapter 6, we present the modeling and the performance comparison of optics, Cu, and CNT-based interconnects for on-chip global connectivity applications.

Interconnect power dissipation is a function of many parameters, and is difficult to properly model its effects at the system level. Several discussions on the performance of electrical and optical interconnect including their comparisons in the off-chip regime have been published [5]-[11]. The comparison discussions, although a good starting point, are either relatively qualitative in nature [9], or are somewhat simplistic as they ignore the role of end-devices [5][6]. Further, the complexity and sophistication of state-of-the-art electrical links is not considered in these comparisons. In this dissertation, we take a more comprehensive view of both Cu and optical systems for short distance, off-chip, bandwidth-sensitive applications. Our primary objective is to compare power dissipation with respect to the relevant parameters such as bandwidth, interconnect length, and bit error rate (BER) by capturing the essential complexity of the two systems, in turn, to optimize the end-devices. We start in Chapter 2 by modeling the high-speed electrical and optical interconnect for short distance off-chip applications, including power optimization method for optical interconnects. We also present a power comparison between electrical and optical interconnects in terms of the critical length defined as the length beyond which optics becomes more power efficient. This length is characterized as a

function of various system requirements (data rate and bit error rate), device parameters (end-device capacitances and coupling loss), and technology.

In Chapter 3, we examine two competing transmitter technologies for optical links, the vertical cavity surface emitting laser (VCSEL) and the quantum well modulator (QWM). There are many tradeoffs between these choices in terms of performance, integration complexity, and reliability. VCSELs simplify packaging by allowing the optical power to be generated on-chip through hybrid bonding and also can provide a larger contrast ratio than modulators. The disadvantage stems from a poor reliability, especially in the high temperature environment of a modern integrated circuit. Modulator based links, especially using the quantum well modulators (QWM), on the other hand, can be more reliable, can be integrated monolithically in silicon, and do not dissipate large power, at least in the transmitter. However, their low contrast ratio, in particular, at the scaled CMOS supply voltages renders a higher power dissipation at the receiver-end. In the first part of this chapter, we simply assume the leading performance of both transmitters, in return, this result gives an idea of the parameter choice (data rate and end-device capacitances). In the second part, we examine the impact of the modulator metrics (insertion loss and contrast ratio) in terms of pre-bias voltage and absorption ratio of “on” and “off” states on the even length defined as the length beyond which VCSEL becomes more power efficient.

The Chapter 4 deals with optimization of the modulator to extend its application to higher bandwidth and shorter distance interconnects. For high-speed optical links to be competitive enough to replace electrical links for short distances, it is imperative to develop methodologies to minimize their power dissipation including design parameters (number of quantum wells, capacitances) and system parameters (bit rate, pre-bias and swing voltages, link efficiency, and input laser power). In this chapter, we extract the modulator metrics from the physical design. The semi-

empirical model for the field dependence of absorption edge [12], dictates the physical design of the modulator. We present an optimized design and system parameters, and discuss the compatibility of the required voltage swing with the scaled CMOS supply voltage.

Chapter 5 deals with the design of a prototype electrical and modulator-based optical link. For electrical links, we implement bipolar current mode simultaneously bi-directional signaling with pre-emphasis at the transmitter, improving signal integrity. On the other hand, for optical links, a quantum well modulator is used as a transmitter driven by buffer chains with separate supply voltage to improve contrast ratio of the modulator. We also address the design technique adjusting the design parameters to minimize total optical link power. Toward the end of this chapter, we show the modulator design and the process for building optoelectronic chips.

In Chapter 6, we compare CNT-based and optical interconnects with Cu interconnects using primary metrics as well as compound metrics for on-chip global wires. We use a bandwidth per link limited by the clock frequency and FO4 inverter delay and assume that the global clock frequency will not increase dramatically in the future. We divided this by the pitch, latency and power to obtain the aforementioned compound metric - bandwidth density/latency/power metric. The reason this metric is important is because for global communication between blocks or cores, the designers care about the bandwidth density, the latency of the links and how much power they need to expend to obtain these links. We extensively examined the impact of device parameters-modulator and detector capacitances for optics, material parameters-mean free path and packing density for CNTs, and system parameters-global clock frequency and switching activity on this metric. Finally, in Chapter 7, we draw the conclusions.

1.4 References

[1] M. T. Bohr, “Interconnect Scaling-The Real Limiter to High Performance ULSI,” *IEDM Technology Digest*, pp. 241-244, 1995.

[2] K. C. Saraswat and F. Mohammadi, “Effect of Interconnect Scaling on Time Delay of VLSI Circuits,” *IEEE Transactions on Electron Devices*, vol. 29, pp. 645-650, 1982.

[3] D. A. B. Miller, “Physical Reasons for Optical Interconnection,” *Special Issue on Smart Pixels, International Journal of Optoelectronics*, vol. 11, no. 3, pp. 155-168, 1997.

[4] Y. Li, E. Towe, and M. W. Haney, “Special issue on optical interconnections for digital systems,” *Proceedings of the IEEE*, vol. 88, no. 6, pp. 723-727, June 2000.

[5] A. V. Mule, A. Naeemi, E. N. Glytsis, T. K. Gaylord, and J. D. Meindl, “Towards a comparison between chip-level optical interconnection and board-level exterconnection,” *International Interconnect Technology Conference*, pp. 92-94, June 2002.

[6] International Technology Roadmap for Semiconductors (2003), *http://public.itrs.net*.

[7] A. Naeemi, R. Venkatesan, and J. D. Meindl, “Optimal Global Interconnects for GSI,” *Transactions on Electron Devices*, vol. 50, no. 4, 2003.

[8] A. Naeemi, A. V. Mule, and J. D. Meindl, “Partition length between board-level electrical and optical interconnects,” *International Interconnect Technology Conference*, pp. 230-232, June 2003.

[9] D. A. B. Miller, “Rationale and challenges for optical interconnects to electronic chips,” *Proceedings of the IEEE*, vol. 88, no. 6, pp.728-749, 2000

[10] M. R. Feldman, S. C. Esener, C. C. Guest, and S. H. Lee, “Comparison between optical and electrical interconnects based on power and speed considerations,” *Applied Opt.*, vol. 27, no. 9, pp. 1742-1751, May 1988.

[11] M. Yoneyama, K. Takahata, T. Otsuji, and Y. Akazawa, “Analysis and application of a novel model for estimating power dissipation of optical interconnections as a function of transmission bit error rate,” *Journal of Lightwave Technology*, vol. 14, no. 1, pp. 13-22, 1996.

[12] D. T. Neilson, “Optimization and tolerance analysis QCSE modulator and detectors,” *IEEE Journal of Quantum Electronics*, vol. 33, no. 37, pp. 1094-1193, July 1997.

Chapter 1: Introduction

Chapter 2

Performance Comparison: Off-chip

2.1 Introduction

Different classes of digital systems impose specific requirements on the communication medium. These requirements pertain to the communication length scale and the figure of merit of relevance (bandwidth or latency). The choice of the communication medium is heavily dependent on these factors. For example, long-haul systems ubiquitously use optical fibers because of low attenuation at high bandwidths. Systems at shorter length scales have traditionally used copper (Cu) interconnects for both latency and bandwidth sensitive applications. However, as the computational bandwidth of the modern integrated circuits (ICs) (measured by the product of the number of transistors and the clock frequency) increases dramatically according to the Moore's law, Cu traces at short distances are struggling to keep up at least in bandwidth sensitive applications, rendering communication bandwidth a bottleneck. This presents a fertile ground for optical medium of communication to penetrate the short distance world, albeit with very different constraints compared to long-haul

communication.

Optical interconnects with low signal attenuation and crosstalk could potentially be very useful in short distance, bandwidth sensitive applications. Already many different backplane optical media have been demonstrated and/or their prospect discussed including polymer waveguides [1], fiber image guides (FIGs) [2][3], fiber ribbons [4], and free space optical interconnects (FSOI) using lens and mirror system [5]-[8].

Several discussions on the performance of electrical and optical interconnect including their comparisons in the off-chip regime have been published [9]-[13]. The comparison discussions, although a good starting point, are either relatively qualitative in nature [11], or are somewhat simplistic as they ignore the role of end-devices [9][10]. Further, the complexity and sophistication of the state-of-the-art electrical links is not considered in the comparisons. In this chapter, we take a more comprehensive view of both Cu and optical systems for short distance, off-chip, bandwidth-sensitive applications. Our primary objective is to compare power dissipation with respect to the relevant parameters such as bandwidth, interconnect length, and bit error rate (BER) by capturing the essential complexity of the two systems. To accomplish this goal we first optimize the optical system by design for power. On the electrical side, we choose a sophisticated, state-of-the-art interconnect for a fair comparison. We model the electrical interconnect attenuation (analytically and using SPICE) including the package effects as well as model various noise sources end-to-end in both systems. In the process, we also identify critical device/system parameters that have the maximum impact on power dissipation in each type of interconnect, while quantifying the severity of their impact. For optical interconnects, these parameters include detector capacitance, coupling efficiency and modulator type, while for electrical interconnect they are receiver sensitivity/offset and impedance

mismatch.

The rest of this chapter is organized as follows. In section 2.2, we present a power-optimizing method for optical interconnects and quantify the system's performance at future nodes. In section 2.3, we tackle the issues related to power modeling in electrical interconnects. We assume a system, which uses simultaneous bi-directional signaling with transmitter equalization and on-chip noise cancellation. Such interconnects are extremely effective in pin-limited systems. Section 2.4 contains power comparison between electrical and optical interconnects in terms of the critical length defined as the length beyond which optics becomes more power efficient. This length is characterized as a function of various system requirements (data rate, BER). Section 2.5 presents the impact of technology on the comparison. Finally, we summarized in section 2.6.

2.2 Optical Interconnect Power Dissipation

The optical interconnect power consists of the transmitter and the receiver powers. In this work, we assume an off-chip laser source at 1.3μm wavelength providing light to silicon CMOS driven modulators. This scheme is an attractive alternative to directly modulating hybrid-integrated Vertical Cavity Surface Emitting Lasers (VCSEL). It enables transmitter power reduction on the CMOS chip by allowing us to consider only the modulator power dissipation, which is much less than VCSEL power dissipation. It also mitigates reliability concerns of lasers in a harsh high-temperature CMOS environment. Finally, it provides the possibility of monolithic integration with Si-CMOS, as it is conceivable to build modulators in Silicon (Si) and/or Si compatible material such as Germanium (Ge). The drawback of

this scheme is the relatively low contrast ratio of the modulators, which can increase receiver power dissipation or in the worst case lead to insufficient BER. The choice of the wavelength not only gives a larger number of photons for a given optical power compared to 850nm (hence larger detector responsivity) but also allows the possibility of monolithic integration of Ge photo-detectors directly onto silicon substrate without the danger of noise in silicon circuits. Although, monolithic integration possibility for both detectors and modulators exists with this scheme, our interconnect performance trends in this work are more representative of the Indium Phosphide (InP) based devices that are hybrid-bonded to Si-CMOS. These devices are reversed biased p-i-n quantum well detectors and quantum well modulators. The hybrid-bonded techniques can already be pushed to yield detector capacitance as low as 50fF including the bond pad and solder capacitances [14][15]. The schematic of the analyzed high-speed optical interconnect is shown in Fig.2.1. The optical medium could be any of the aforementioned possibilities. Its relevance on power calculation is captured through its attenuation and coupling efficiency from/to end-devices.

2.2.1 Modulator Power Dissipation

Both the dynamic and the absorption related, static components of the modulator power dissipation are considered. Modulator is driven by an exponentially sized buffer chain to minimize delay and increase speed. Thus, the dynamic power includes the capacitance of both the modulator and the buffer chain. The static power dissipation depends on the absorbed optical power in the “on” and the “off” state. An ideal modulator should have zero insertion loss *(IL)* (optical power absorbed during the “on” state) and infinite contrast ratio *(CR)* (ratio of modulator output optical power in the “on” and the “off” states). However, all modulators exhibit a non-zero *IL* and

only a finite *CR.* We use the *CR* and *IL* to calculate the current in each of the binary state. Next, we multiply it by respective voltage biases and take the average to get power dissipation [14].

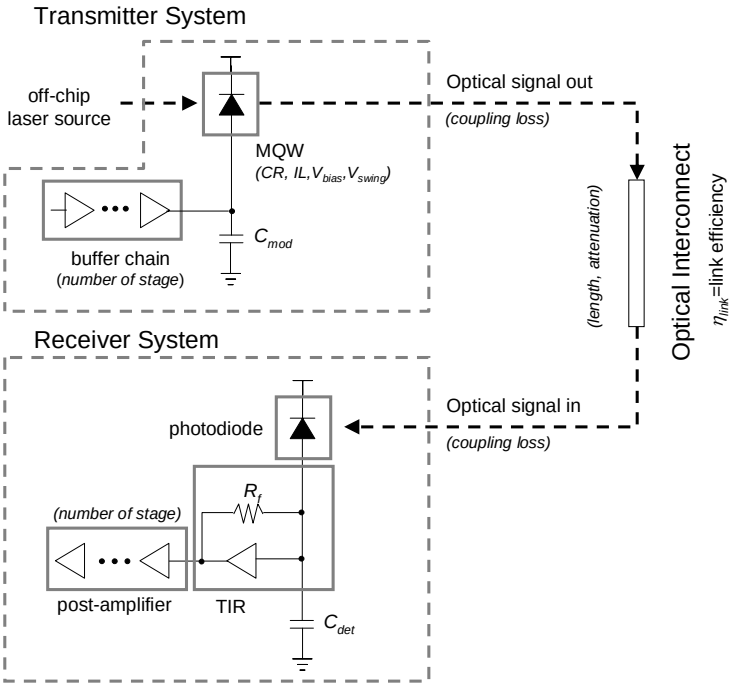


Fig.2.1 Schematic showing board-level high-speed optical interconnect

Thus, for a one to one transmitter/receiver pair, we end up with

$$P_{static\,\mathrm{mod}ulator} = \frac{P_{opt,rec}}{\eta}\frac{q\lambda}{h\nu}\left(\frac{V_{bias}(1+IL-\frac{1-IL}{CR})-V_{dd}IL}{(1-IL)(1-\frac{1}{CR})}\right) \qquad (2.1)$$

Here, $P_{opt,rec}$ is the average optical power at the receiver, ν is the velocity of light in

the waveguide medium, λ is the wavelength of the laser source, and V_{bias} is the DC bias applied to the modulator in the highly absorbing state. V_{dd} is the voltage swing (supply voltage of the CMOS generation). In the less absorbing ("on") state, the modulator is driven to a lower voltage of V_{bias}-V_{dd}. η, the optical power transfer efficiency, is the optical power at the receiver divided by that at the output of the modulator. The efficiency is usually less than one due to both the coupling losses at the transmitter and the receiver ends as well as the losses incurred by the optical transmission medium. In this work, the plots in which we have explicitly used interconnect length as an independent variable, correspond to the waveguide medium. Here, the loss was calculated to be about 0.082 dB/cm at wavelength of 1.3μm using theoretical analysis in [16] and the published values in [14]. As is obvious from (2.1), we can lower the static modulator power by operating at near zero bias voltage in the highly absorbing state ("0") and V_{dd} bias in the less absorbing state ("1"). Modulators exhibiting properties close to this ideal modulator can be realized by using resonant cavity around QWM. At low bias, they exhibit absorption such that the front and the effective back mirror reflectivities are the same, leading to maximum passes in the cavity and high total absorption. Increasing the bias, although increases individual QW absorption, but decreases the effective back mirror reflectivity leading to an asymmetric cavity with less passes, hence, lesser total absorption. We will refer to an ideal modulator along these lines as modulator 1 and the commonly used reflective mode modulator as modulator 2. The bias, *CR* and *IL* values for modulator 2 are taken from [14]. An optimization of *IL* and *CR* in modulators to further minimize total interconnect power dissipation is possible [17], which is discussed in Chapter 3.

2.2.2 Receiver Power Dissipation

The optical receiver is assumed to be the photo-detector (responsivity ~ 0.5A/W) followed by non-integrating transimpedance amplifier and gain stages [18]. Its design and power dissipation is detailed in an earlier work [19] and summarized in Appendix A. The analytical design included establishing the width, the feedback resistance of the front-end, and the number of subsequent gain stages at a given input optical power (IOP) and detector capacitance. It was constrained by bandwidth, BER through receiver noise, and supply level output swing requirements. The transistor related parameters were taken from the International Technology Roadmap for Semiconductors (ITRS) [20]. The power was found to reduce with higher IOP and lower detector capacitance. Building upon that work, we have now verified the design at 180nm technology node using BSIM3v3 technology[1] SPICE simulations. As an example, we show the SPICE generated eye-diagram of an analytically designed 4Gb/s receiver with 20μW reflectivity difference (*RD*; receiver optical power difference between the on and the off states) and 100fF detector capacitance (Table 2.1 and Fig.2.2). Although, the noise is not simulated here, the clean eye indicates sufficient bandwidth.

Table 2.1 also shows a reasonable agreement between power dissipation obtained using SPICE and analytically calculated values. A small difference is accounted by two factors. Firstly, the SPICE simulations required an additional gain stage (5 stages) compared to the analytical model (4 stages). This is because in the model we assume the post-amplifier gain to be the DC gain (product of transconductance and the output resistance). This turns out to be a slight overestimate at high bit rates, hence an additional stage. Secondly, we only consider static power dissipation at each stage in the analytical model. This is a good approximation at low

[1] University of California, Berkeley Device Group, CA, USA [Online]. Available: www-device.eecs.berkeley.edu/~ptm/download.html

swings for most of the gain stages, but toward the last stage as signal becomes large, the static power is reduced. However, this is compensated by the increase in the dynamic power dissipation. The favorable comparison with SPICE simulations lends confidence in our analytical models, especially for future technology nodes where SPICE technology files may not exist.

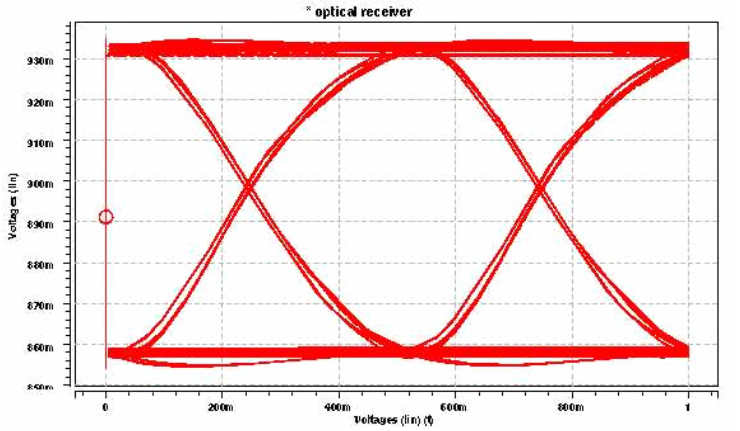


Fig.2.2 Eye diagram at the output of transimpedance receiver (4Gb/s, *RD*=20μW, C_{det}=100fF, SPICE simulation with 180nm BSIM3v3)

Table 2.1 Example of optimized design parameters and power dissipation

Analytically obtained Optimized design parameters	
Front-end transistor size	94λ (2λ=180nm technology node)
Feedback resistance	375 Ω
Number of post-amplifier stage	4
Power dissipation (bit rate=4Gb/s)	
Analytical modeling	22.86mW
SPICE simulation (number of post amplifiers)	18.10mW (4 stages) 21.02mW (5 stages)

2.2.3 Power Dissipation Minimization

Fig.2.3 illustrates our optical interconnect power minimization methodology. The increase in the optical power increases the modulator power but decreases the receiver power as discussed above. This lends to an optimal laser power at which total interconnect power (receiver and modulator) is minimized. The figure demonstrates the minimization for two different losses and for modulators 1 and 2. The receiver power does not change with laser power on continuous bases beyond a certain point as it goes into gain-limited regime [17]. As expected, modulator 2 yields larger power dissipation than modulator 1. Also, the receiver power is dominant over the modulator power. A higher loss (6dB in Fig.2.3) in optical power through either less efficient coupling and/or greater attenuation in the optical interconnect (longer lengths), results in a larger receiver power dissipation. This is due to a lower reflectivity difference between the on and the off states at the receiver with larger optical power loss, which is tantamount to Digital Signal to Noise Ratio (DSNR) deterioration. An increase in receiver power, in turn, results in a larger optimal laser power and total power dissipation. The numerator of the DSNR is the product of the reflectivity difference and responsivity of the photo-detector (R_s) and is given by

$$I_{on} - I_{off} = R_s(RD) = R_s(P_{on} - P_{off}) = R_s \eta P_l (1 - IL)(1 - \frac{1}{CR}) \tag{2.2}$$

where, I_{on} (P_{on}) and I_{off} (P_{off}) is the photo-detector current (receiver optical power) in the on and the off states, respectively and P_l is the laser power. Thus, in the high loss (low RD) case, the noise in the receiver has to be reduced to maintain same DSNR. This costs power dissipation.

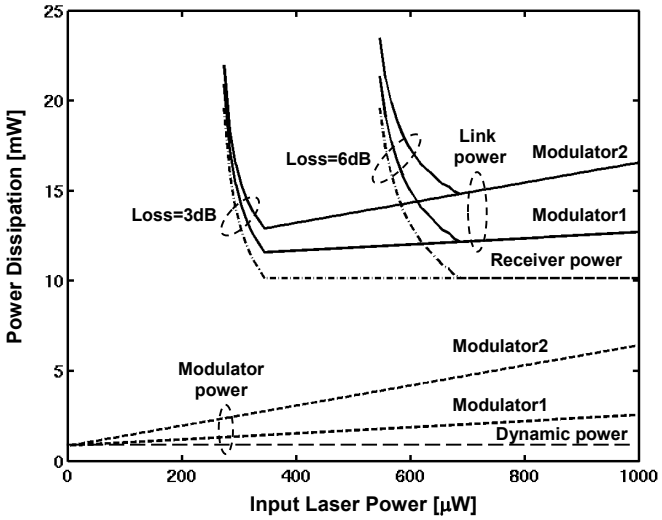


Fig.2.3 Power optimization method for optical interconnect (C_{det}=250fF, BR=6Gb/s)

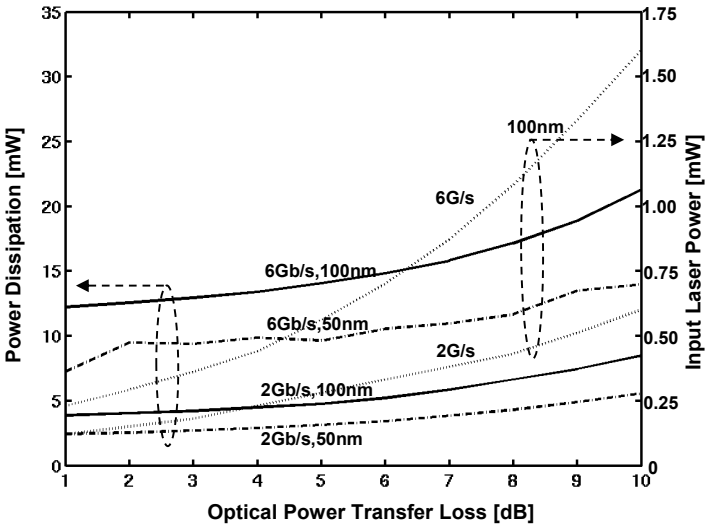


Fig.2.4 Optical interconnect power dissipation and optimized input laser power for 100nm/50nm technology nodes with bandwidths (C_{det}=250fF for Modulator2)

Fig.2.4 exhibits both the optimum laser power and the resulting minimum power dissipation as a function of loss for two different bit rates (2 and 6 Gb/s). Increase in the power dissipation with bit rate is almost entirely due to a larger power dissipation in the receiver at higher bit rates. Fig.2.4 also shows the reduction in power dissipation with technology scaling (100nm and 50nm) due to improvement in receiver transistors. The detector capacitance of 250fF in this plot is somewhat pessimistic. Capacitance approaching 50fF has been demonstrated [15]. We will later quantify the impact of lowering the detector capacitance on power dissipation.

2.3 Electrical Interconnect Power Dissipation

For the electrical interconnect, we choose a link capable of simultaneous bi-directional signaling. This scheme, by enabling full-duplex channels, provides higher aggregate bandwidth over smaller number of pins, a scenario particularly useful in pin-limited chips. However, it uses a more complicated detection scheme, where a transmitter replica is fed as a reference to the differential amplifier receiver to isolate the received and the transmitted signal (Fig.2.5). We also choose low swing current mode, bipolar, differential signaling scheme. The rationale for these choices was driven by the achievement of maximum noise immunity. Differential signaling dramatically reduces signal return crosstalk and facilitates a noise-free receiver reference [21]. High impedance of current mode signaling, further, enables noise immunity to power supply [21].

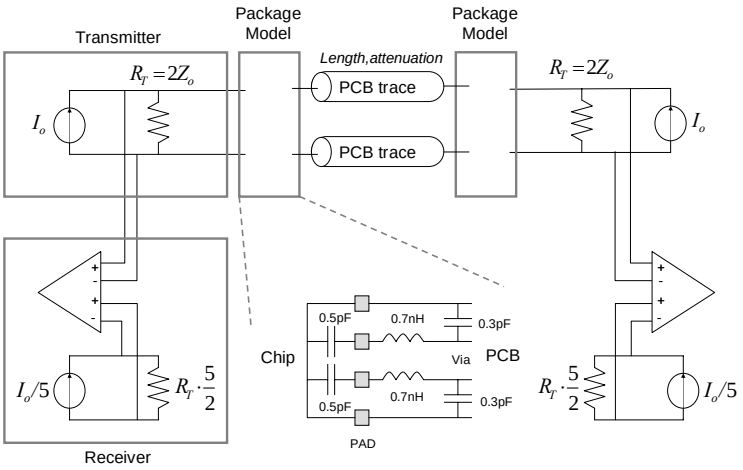
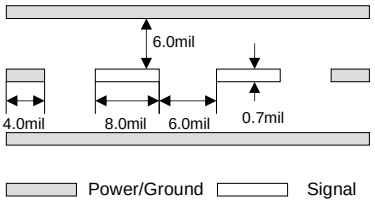


Fig.2.5 Schematic showing board-level electrical interconnects

L	302 [nH/m]
C	148 [pF/m]
Z_o	45 [Ω]
R_o	4.71 [Ω/m]
R_s	$1.313\times10^{-3}[\Omega/m\sqrt{Hz}]$
G_o	0
G_d	$9.929\times10^{-12}[\Omega/m\cdot Hz]$



$R = R_o + R_s\sqrt{f}$ (Skin Effect)

$G = G_o + G_d f$ (Dielectric Loss)

Fig.2.6 Summaries of the board trace parameters and the corresponding SPICE parameters used for dielectric and skin effect loss

The board trace dimensions are chosen to yield characteristic impedance, Z_0, of 45Ω (Fig.2.6) on a high performance GETEK board. This board, although expensive compared to usual FR4 board, provides lower dielectric loss, hence lower signal attenuation (loss tangent is 0.01). Striplines, as opposed to microstrips, are used to eliminate forward cross-talk. Flip-chip package model with lower parasitics inductance and capacitance is chosen (Fig.2.5, [21]). Complete ISI cancellation is assumed using a transmitter side pre-emphasis equalization with multi-tap FIR filter

[22]. The rise time is taken to be a third of the bit period. This is assumed to provide a reasonable compromise between smaller rise time requirement for adequate timing margin and larger rise time for lower noise. An on-chip cancellation circuit to reduce reverse channel crosstalk due to package reflections is also assumed. The principle behind this concept is the following. Since we know the exact value and the time at which reflection from package parasitics (and connectors, if applicable) arrive back at the receiver, we can synchronously generate the same voltage using an additional source and feed it into the reference of the receiving differential amplifier to partially cancel it. In short, full consideration was given to maximize electrical interconnect performance with sophisticated schemes for fair comparison with its optical counterpart.

We consider the power dissipated in the termination resistors related to current swing requirement, as this power is becoming increasing fraction of the total power [23]. Also, this power critically depends on the attenuation and noise characteristics of interconnects, attributes where there is a stark difference between electrical and optical media. The idea is to model the attenuation and noise sources in the electrical interconnect as a function of the bit rate and length. From this, we backtrack the minimum current swing required at the transmitter for an adequate signal to noise ratio at the receiver. Finally, from current swing we calculate the power dissipation.

The BER in an electrical interconnect system is approximated by [21]

$$BER = e^{-\frac{VSNR^2}{2}} \text{ where } VSNR = \frac{V_{nm}}{V_{gaussian}} \tag{2.3}$$

Here, *VSNR* is the voltage signal to noise ratio. V_{nm}, the net noise margin, is given by the difference of half the signal swing and the sum off all worst-case noise sources at

the receiver. $V_{guassian}$ is the standard deviation of all the statistical noise sources, which are assumed to be uncorrelated. From (2.3), we have the condition on the net required noise margin for adequate BER

$$V_{nmreq} = V_{guassian}\sqrt{2ln(\frac{1}{BER})} \tag{2.4}$$

The net available noise margin at the receiver depends on two factors: the attenuated signal swing and the sum of all worst-case noise sources. The attenuation in signal swing is modeled extensively and will be described subsequently. The worst-case noise sources are of two kinds: Proportional to signal swing or independent of it (fixed sources). The first type of proportional noise source is attenuated by the trace just as the signal because it is acquired at the transmitter end. Its proportionality constant is denoted by K_A. This includes trace crosstalk, impedance mismatch and package reflections. The second type of proportional noise source is acquired at the receiver-end, hence is not attenuated by the board trace (denoted by K_U). K_U is present only in the case of simultaneous bi-directional signaling due to the opposite direction transmitter at the receiver end. It includes reverse channel crosstalk, package reflections and transmitter replica mismatch. Finally, the fixed noise sources (V_{NF}) arise due to the receiver offset and its sensitivity.

The effect of transmitter-end pre-equalization with multi-tap filters is to increase the voltage fraction from *1-2A* to *A*, where A is the attenuated fraction of the signal at the receiver at a particular bit-rate [21]. The post-equalization swing at the receiver is, then, *A* times the swing at the transmitter ($V_{swtrans}$), resulting in a gross noise margin of half this value. To meet the BER, the available net noise margin should be greater than the required net noise margin, hence,

$$V_{swtrans}\left(\frac{A}{2}-AK_A-K_U\right)-V_{NF}>V_{nmreq} \tag{2.5}$$

Thus, from (2.4) and (2.5), the minimum swing required is

$$V_{swtrans}=2\frac{V_{nmreq}+V_{NF}}{A(1-2K_A)-2K_U}=2\frac{V_{NF}+V_{guassian}\sqrt{2ln(\frac{1}{BER})}}{A(1-2K_A)-2K_U} \tag{2.6}$$

For differential, bipolar current mode signaling with parallel termination ($2Z_0$), and a current swing from $-I_0$ to $+I_0$. **(**Fig.2.5**)**, the transmitter side voltage swing is given by $2(I_0/2)\,(2Z_0)=2I_0Z_0$. Thus, the required I_0 (one way swing) is given by

$$I_0=\frac{V_{NF}+V_{guassian}\sqrt{2ln(\frac{1}{BER})}}{Z_0[\,A(1-2K_A)-2K_U\,]} \tag{2.7}$$

The component of total power which is dissipated in the termination resistance for one-way signaling is the sum of the power dissipated in the two termination resistances and the power dissipated in the replica transmitter circuit to cancel the opposite side transmitter signal. For further power minimization, we used a scaled current (factor of 5) and increased impedance by the same factor in the replica circuit. Thus,

$$P_{term}=\left(\frac{I_0}{2}\right)^2(2Z_0)+\left(\frac{I_0}{2}\right)^2(2Z_0)+\left(\frac{I_0}{5}\right)^2(5Z_0)=1.2I_0^2Z_0 \tag{2.8}$$

where I_0 is given by (2.7). The other sources of power dissipation in this link are the

transmitter and receiver logic circuit power, equalization power, and the power due to additional transmitter for canceling the near-end LC tank package reflections. The transmitter logic power includes the dynamic power due to exponentially sized buffer chain. For low-end receiver logic circuit, we estimate the tail current of the differential amplifier to be about 100μA. A two-stage amplifier was needed to amplify the input signal to the power rails. The power dissipation in the high-end receiver with very low sensitivity and offset is assumed to be the same, which is optimistic. Amongst the sources considered, the termination resistance power is found to be dominant. The transmitter logic as well as the cancellation circuit power expectedly tracks the power in the termination resistance. Equalization power is neglected, as subsequent taps are scaled version of the main transmitter current. The power due to clock and timing circuits for clock recovery is not considered in this work. These components are also omitted in the case of optical interconnects for fair comparisons.

Table 2.2 summarizes noise sources in electrical interconnect assuming 5% mismatch between termination resistances and the characteristic impedance of the printed circuit boards (PCB) trace. We have ignored the noise due to connector crosstalk, which will make the electrical interconnects power results slightly optimistic. The reverse crosstalk was estimated with SPICE simulations. The Gaussian noise was assumed to be 5mV [24]. We considered two types of electrical receivers for each bit rate. The high-end receiver has an offset of 8mV and sensitivity of 0.8mV [25]. When using PCB for multi-gigabit data rate, attenuation due to both the skin effect loss and dielectric loss become extremely important and it is imperative to model it accurately.

Table 2.2 Summary of noise sources in electrical interconnects

<table>
<tr><td colspan="6">Proportional noise</td></tr>
<tr><td colspan="5">Attenuation (1-A for PCB Length=50cm, 6Gb/s)</td><td>0.239</td></tr>
<tr><td colspan="6">Noise sources due to unidirectional signaling (Attenuated, K_A)</td></tr>
<tr><td colspan="2">Near-end cross-talk</td><td colspan="2">Impedance mismatch</td><td>Transmitter impedance mismatch</td><td>Package (LC)</td></tr>
<tr><td colspan="2">0.0005</td><td colspan="2">0.025</td><td>0.025</td><td>0.075</td></tr>
<tr><td colspan="6">Noise sources due to bidirectional signaling (Un-attenuated, K_U)</td></tr>
<tr><td colspan="2" rowspan="2">Reverse-channel crosstalk</td><td colspan="3">Reverse crosstalk (same interconnect)</td><td rowspan="2">Transmitter offset</td></tr>
<tr><td colspan="2">Impedance mismatch</td><td>On chip cancellation</td></tr>
<tr><td colspan="2">0.022</td><td colspan="2">0.01</td><td>0.05</td><td>0.05</td></tr>
<tr><td colspan="6">Fixed noise (V_{NF}) (Un-attenuated)</td></tr>
<tr><td colspan="3">Gaussian noise ($V_{gaussian}$)</td><td colspan="3">5mV</td></tr>
<tr><td>Low-end</td><td rowspan="2">Receiver offset</td><td>17.4mV</td><td rowspan="2">Receiver sensitivity</td><td>20mV</td></tr>
<tr><td>High-end</td><td>8.0mV</td><td>0.8mV</td></tr>
</table>

Attenuated/remaining fraction of the signal at the receiver due to these two effects is given by

$$A = e^{-R/2Z_o} e^{-GZ_o/2} \tag{2.9}$$

$$R = \frac{R_{DC}}{2}\left(\frac{f}{f_s}\right)^{1/2}, \; G = 2\pi f C \tan\delta_D \tag{2.10}$$

where, R_{DC} is constant DC resistance, f_s is the frequency, where the skin depth is equal to the height of the conductor, and $\tan\delta_D$ is the loss tangent. We have calculated attenuation for our dimensions using the comprehensive analytical model (accounts for frequency dependent loss tangent) developed in [26] (analytical model) and have

compared the results with SPICE simulations (including package effect). The SPICE models use a constant loss tangent. As shown in Fig.2.7, the frequency dependency in the loss tangent causes more attenuation at higher bit rates in the analytical model compared to SPICE without package effect. However, including package effect with SPICE increases attenuation above the analytical model. The figure also decouples the skin-effect and the dielectric loss and clearly shows that dielectric loss becomes more limiting at high frequency.

An interesting point that follows from equation (2.6) is that there exists a minimum allowed attenuated signal (maximum attenuation) for simultaneous bi-directional signaling, beyond which, the denominator in (2.6) becomes negative and noise margin will never be met. This limit is given by

$$A_{min} = \frac{2K_U}{1-2K_A} \tag{2.11}$$

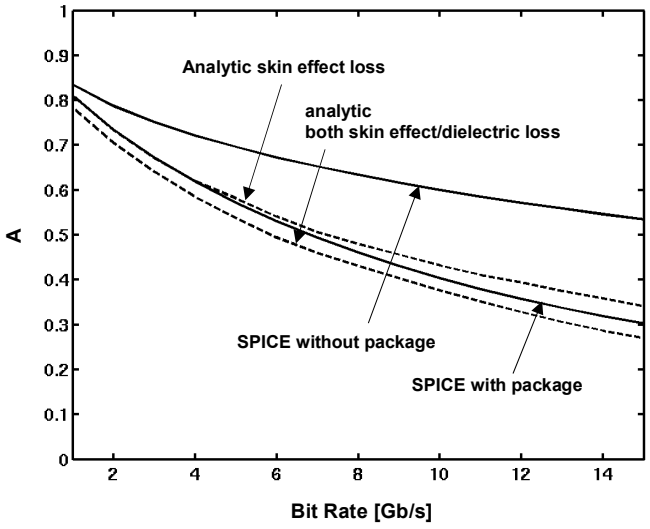


Fig.2.7 Attenuation comparison between SPICE and modeling with skin effect, dielectric loss, and package effect

Since attenuated signal limit in (2.11) depends on both bit rate and length (equation (2.9) and (2.10)), it follows that for a given interconnect length, there is a maximum allowed bit rate and vice versa. This bit rate-length contour is plotted for our case in Fig.2.8. However, the power dissipation will become prohibitively high much before this limit is reached. Fig.2.8 also compares the maximum bit rate limit with other calculations in the literature as well as some experimental results. Reference [27] calculates the maximum bit rate for a single-ended system by using equation (2.9) to ultimately obtain the step response of interconnects. Using this step response and an arbitrarily chosen signal swing requirement at the receiver, one obtains the minimum permissible bit time for signal to rise to the chosen value, hence the maximum bit rate. In practice, the swing requirement would depend on the noise in the system. This yields the area/length^2 limit for electrical interconnects [27]. The maximum bit rate for two different swing requirements (50 and 68% of the transmitter side voltage) is shown. In contrast, our calculation in the figure relies on explicit noise estimation and is for bi-directional signaling.

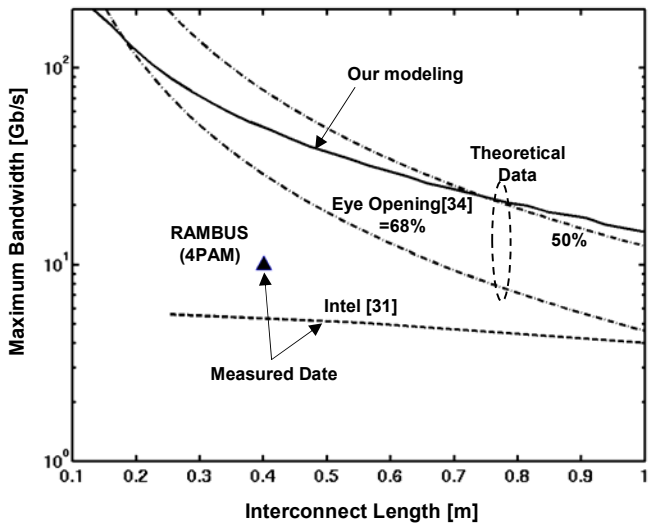


Fig.2.8 Maximum bandwidth of electrical interconnect with simultaneous bi-directional signaling

2.4 Impact of Device and System Parameters on the Critical Length

Fig.2.9 compares the electrical and optical interconnect power vs. length at 4 and 6Gb/s. Electrical interconnect consists of two sets of curves corresponding to different receiver offset/sensitivity. Electrical interconnect power rises with length and bit rate due to a larger attenuation, and a greater impact of un-attenuated as well as fixed noise sources. At higher bit rates a smaller rise time would further increase noise sources such as parasitics reflections from package LC tanks. The optical interconnect power dissipation also rises with length owing to a greater loss in optical power resulting in a smaller η, albeit this rise is slower than that for electrical interconnect. Beyond a critical length, optical interconnect yields lower power. This critical length reduces at higher bit rates. The figure also explicitly quantifies the impact of improving fixed noise, V_{NF} (receiver sensitivity/offset) on electrical interconnect power dissipation.

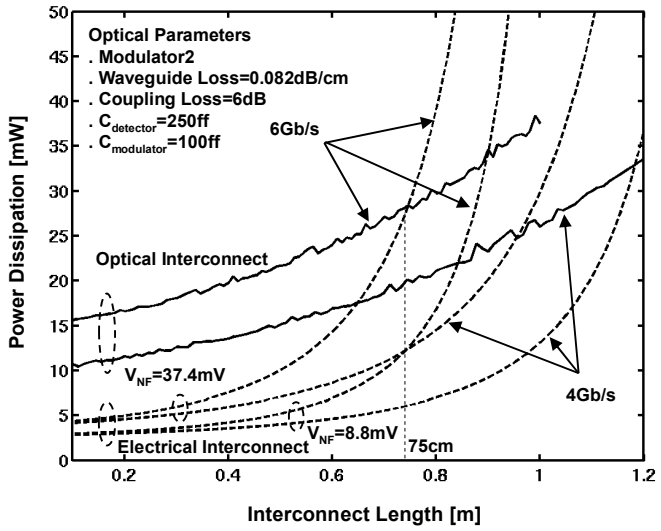


Fig.2.9 Power comparison between electrical and optical interconnects for the modulator 2

In Fig.2.10, we explicitly quantify the impact of critical device/system parameters in respective interconnects, on the critical length. Specifically, for optical interconnect, we consider the role of detector/modulator capacitance, coupling loss and ideal modulator 1, whereas, for electrical interconnect, we examine the role of receiver sensitivity/offset on the critical length. The critical length with modulator 1 is found to be 44 and 20 cm when compared with high and low-end electrical receivers, respectively, with 6dB coupling loss and 50fF detector capacitance (low-end with mod 1 not shown in the graph). Both coupling loss and detector capacitance play a pivotal role in dictating critical length. For example, bringing down the detector capacitance from 250 to 50fF with 3dB coupling loss reduces critical length from about 80 to 45cm with modulator 2.

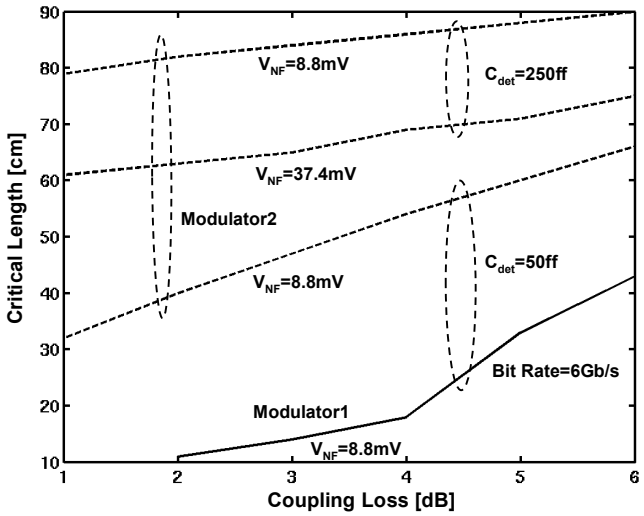


Fig.2.10 Critical length in terms of design parameters

In Fig.2.11, the critical length is shown in terms of bit rate of the system. For modulator 2, this length gradually reduces to about 40cm at 15Gb/s, while the slope is more shallow for modulator 1. These results are for 6dB coupling loss. If the coupling

loss is reduced further to 3dB, the critical lengths could come down by significant amount as shown in Fig.2.11 for 6Gb/s. We see an apparent saturation of critical length at high bit rates. This is because we tend to underestimate electrical interconnect power at high bit rates. With bit rate increase, we account for the impact of worsening trace attenuation on power dissipated in the termination resistances, but neglect the power increase in the electrical transmitter and receiver due to complexity in its modeling. If this power increase was incorporated, the zero length power dissipation in Fig.2.10 would increase with bit rate rather than remaining constant. This would yield a lower crossover point (critical length) between electrical and optical curves than what we have calculated, with the difference accentuated at higher bit rates.

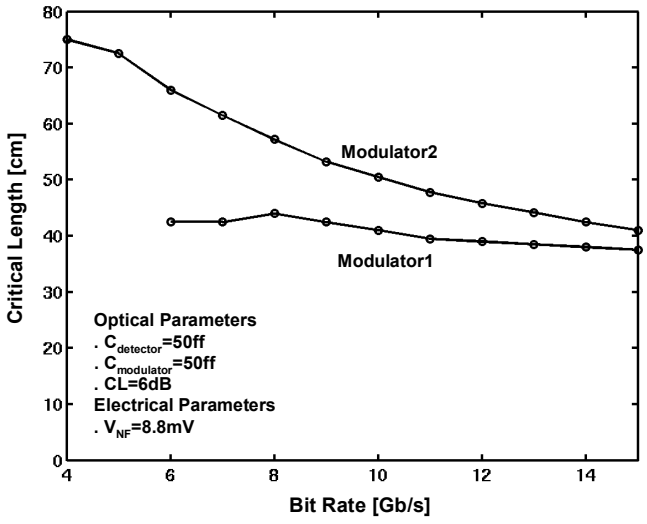


Fig.2.11 Critical length in terms of bit rate

Next, we plot the critical length as a function of BER requirement (implicitly in terms of digital signal to noise ratio, DSNR) in Fig.2.12. Different BER is demanded in different system applications and high BER can be tolerated if explicit

error correction schemes are utilized. For example, in communication application BER between 10^{-15} to 10^{-12} is deemed sufficient, whereas, server systems, particularly if they are not deploying error correction, require BERs less than 10^{-15}. From Fig.2.12**,** it is clear that for small BER values, the critical lengths are smaller and optical interconnects have advantage over electrical interconnects. Thus, optical interconnects are more power-favorable for systems where data reliability criteria is demanding.

Finally, we examine the sensitivity of critical length on the mismatch between termination impedances and the characterization impedance of the PCB trace as this constitutes a significant noise source in electrical interconnects (Fig.2.13). The critical length is found to substantially increase with small reduction in the impedance mismatch.

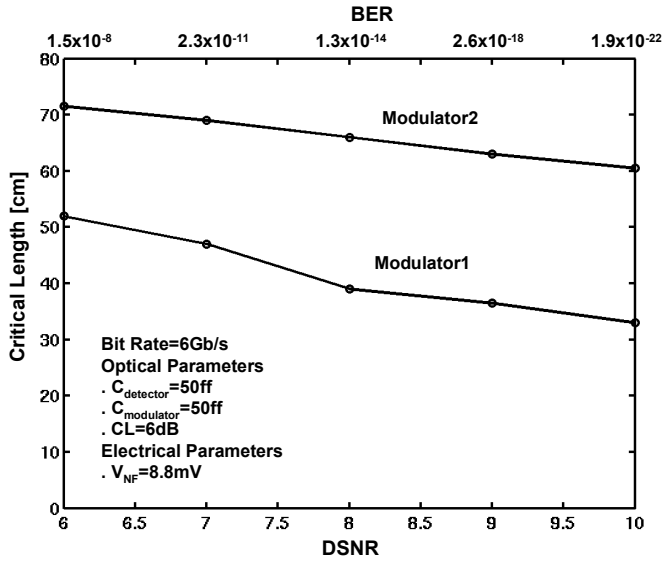


Fig.2.12 Critical length in terms of BER

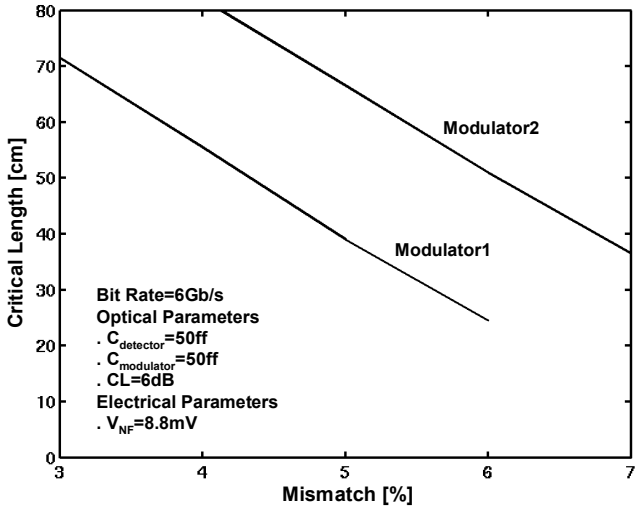


Fig.2.13 Critical length in terms of mismatch in terminator

2.5 Impact of Technology

An explosive growth in on-chip computational bandwidth due to both higher integration levels and faster transistors demands a commensurate increase in off-chip I/O bandwidth. This demand for inter-chip applications is depicted in Fig.2.14 according to the ITRS [20]. For comparison the figure also depicts a more slowly rising transistor performance in the form of “On Current” increase with technology node. The traditional copper (Cu) based board-level interconnects face severe impediments in meeting the high I/O bandwidths due to large skin effect and dielectric losses as discussed in prior sections. The design solutions to overcome these deleterious effects require complex signal processing at the interconnect endpoints, which results in a larger power dissipation and area requirement. Optical interconnects offer a powerful alternative, potentially at a lower power as presented in prior section

2.4, analyzing the role of the end-devices and various systems parameters on the power dissipation of both optical and electrical interconnects, while comparing the two. In this section, we extend this work by combining it with the ITRS to evaluate the impact of technology scaling on the performance of the two types of interconnects for inter-chip applications. Hereafter, we deals with the optical interconnect with the modulator 1 because the critical length becomes close enough for very high bit rate as shown in Fig.2.11, hence this did not impact the general trend presented in this section.

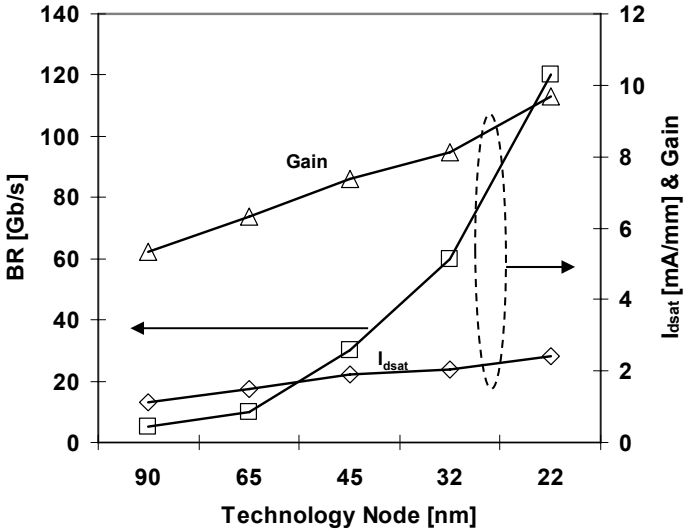
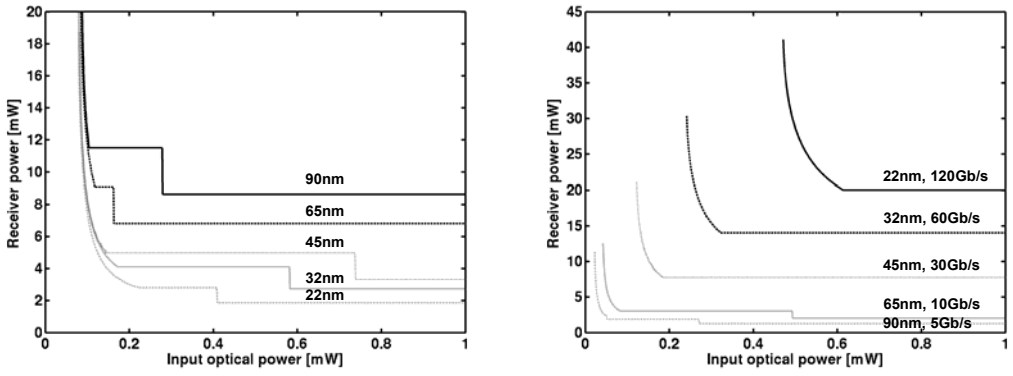


Fig.2.14 ITRS roadmap of required bit rate (BR) of inter-chip communication and transistor performance

As shown in Fig.2.14, technology nodes imply both improved transistor performance and required bit rate. Fig.2.15(a) depicts the receiver power in terms of input optical power with different technology nodes at a fixed bit rate (20Gb/s). In this sense, it decouples the effect of transistor improvement with bit rate, both of which occur with scaling. The receiver power decreases with improved transistor performance, while required bit rate impacts in an opposite way. Fig.2.15(b) includes both of transistor improvement and required bit rate, indicating that the impact of

required bit rate outweighs that of transistor improvement. The input optical power increases with scaling and is about 0.6mW for 22nm technology node. This optical power requires more than 1mW assuming the link loss of 3dB, hence technology scaling demands high power laser source.



(a) (b)

Fig.2.15 Receiver power vs. input optical power in terms of technology node (a) C_{det}=50fF, fixed bit rate=20Gb/s (b) C_{det}=50fF, bit rate of ITRS

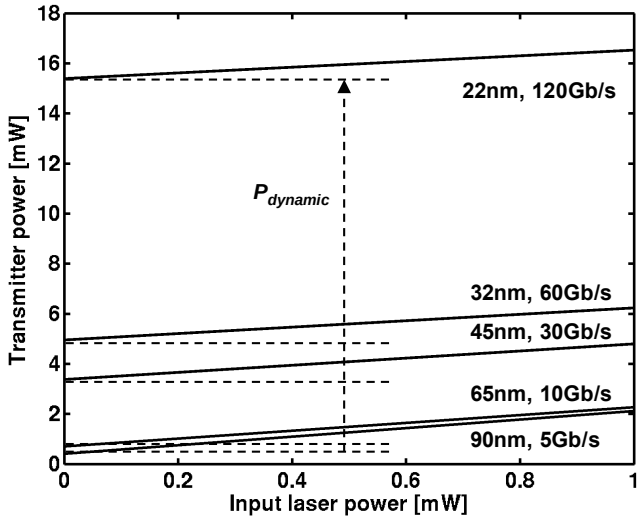


Fig.2.16 Transmitter power vs. input laser power in terms of technology node showing the dramatic increasing dynamic power of the modulator

Fig.2.16 presents the transmitter power in terms of input laser power with different technology node. The static power of the modulator depends on the device parameters, hence relatively constant with scaling. The dynamic power, on the other hand, drastically increases due to increase of bit rate and renders transmitter power comparable to receiver power.

Fig.2.17 compares the electrical and the optical power dissipation for a scaled technology at a fixed bit rate. The detector capacitance is assumed to be 25fF. Both optical and electrical interconnect power dissipation shows marked reduction with technology scaling. However, the critical length (defined as length above which optical interconnects is more power efficient) remains approximately constant. This is because for small detector capacitance (25fF) in the case of optical interconnects, optimized receiver transistor sizes are already small. Hence, a further reduction in the power dissipation (proportional to transistor width) with advanced technology is negligible and is comparable to that of electrical interconnect.

Fig.2.18 explicitly shows the critical length as a function of technology node at various required bit rates. Critical length sharply reduces with bit rates making optics overwhelmingly advantageous at high bit rates. This is primarily due to the debilitating high-frequency effects in the case of electrical interconnects, which compromise the signal integrity and require additional power to satisfy the target bit error rate criteria.

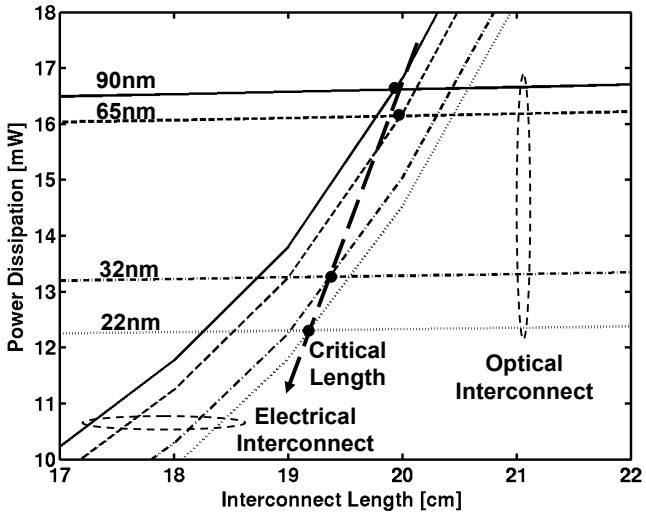


Fig.2.17 Power dissipation vs. interconnect length for BR=30Gb/s and C_{det}=25fF

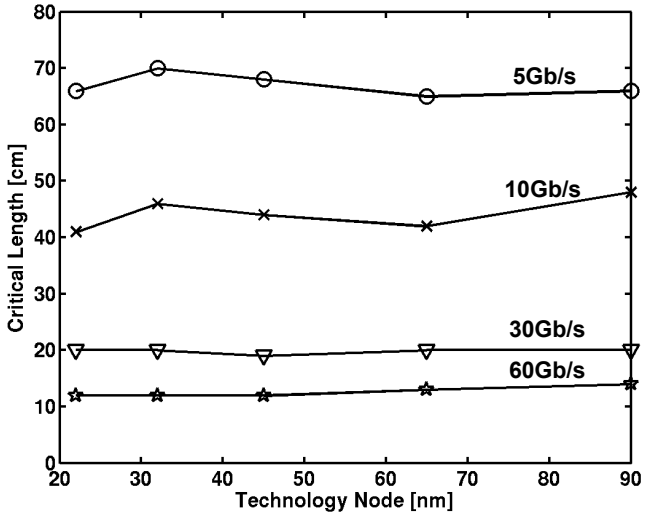


Fig.2.18 Critical length vs. technology node for C_{det}=25fF

Fig.2.19 compares electrical and optical power dissipation, incorporating both transistor performance improvement and higher bit rate demand with technology

scaling. It assumes a detector capacitance of 50fF. Below 10Gb/s corresponding to technology nodes greater than 65nm, the critical length is larger than 50cm, making optics viable for inter-board (backplane) communications. At 45nm technology node, this reduces to about 20cm, making optics suitable for even inter-chip applications. However, beyond 45nm technology node, the required input laser power for optimal optical interconnect becomes quite large. To alleviate this requirement, detector capacitance and total loss (coupling and transmission), must be minimized.

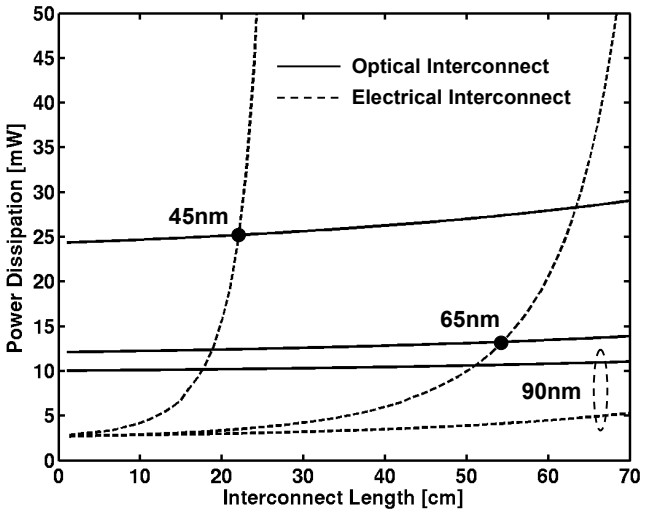


Fig.2.19 Power comparison for C_{det}=50fF

Fig.2.20 is a similar plot to Fig.2.19 but is shown for a smaller detector capacitance of 10fF. With this capacitance not only the critical length is lower, but the optimal required laser power (not shown) beyond 45nm is reasonable, making optics extendible down to 22nm node. The critical length is less than 10cm for 32nm and 22nm technology nodes, which puts optics strongly in contention for inter-chip applications. Interestingly, for 90nm and 65nm technology nodes, optical interconnects yield lower power for all range of interconnects. This is a result of our

assumption of an invariant power dissipation at the electrical receiver with technology scaling. In practice, the receiver power is expected to decrease; however, this effect is not as important when compared with the impact of bit rate and transistor performance because the critical length crossover occurs where the electrical interconnect power increases rapidly.

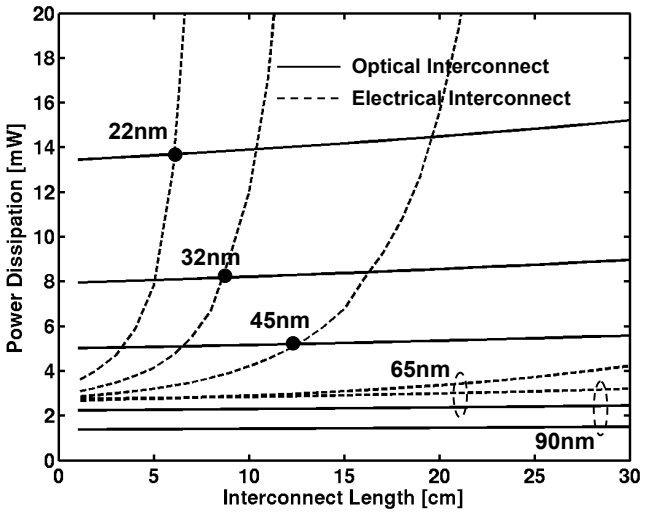


Fig.2.20 Power comparison for C_{det}=10fF

In Fig.2.21, we quantify the critical length vs. technology node for different detector capacitances. The technology scaling incorporates both transistor improvement and higher demand on bit rates. The figure clearly elucidates the importance of lower detector capacitance in facilitating the insertion of optical interconnects. It also shows dramatically that optical interconnects will become more and more favorable in the future by showing a rapid reduction in critical length with technology scaling.

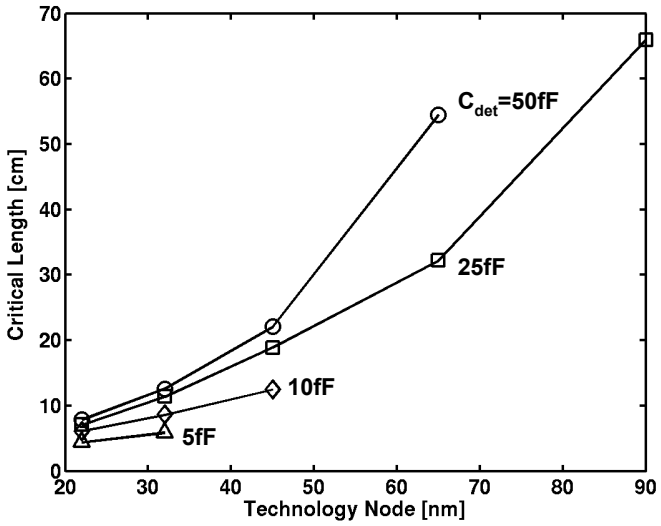


Fig.2.21 Family of curves corresponding to different detector capacitances plotting critical length as a function of technology scaling. Technology scaling incorporates both transistor performance improvement and higher bit rate demand.

2.6 Summary

We have done extensive power dissipation comparison between electrical and optical interconnects for bandwidth sensitive applications in 10cm to 1m range of interconnects. This comparison, among other things, included introduction of a sophisticated power optimization scheme for optical interconnects and detailed noise and attenuation modeling in electrical interconnects. Based on this modeling, power dissipation was calculated as a function of length and bandwidth. We find that beyond a critical length, within the application range, power optimized optical interconnects dissipate lower power compared to the state-of-the-art high-speed electrical signaling scheme. We have further quantified the impact of various device and system

components in electrical and optical interconnects on the critical length. This falls under three categories. 1. On the optical side, we have explicitly quantified the impact of detector/modulator capacitance, coupling loss and modulator type on the critical length. Whereas, it is also implicitly possible to conclude from equation (2.2) that the impact of detector responsivity would be similar to that of coupling loss. 2. On the electrical side, we have characterized critical length as a function of receiver sensitivity/offset and impedance mismatch. 3. On the system demand side, we have studied the critical length as a function of bandwidth and BER. This gives both optical device designers as well as electrical circuit designers a framework to assess the system's level (power) impact of various figures of merit of the devices. When compared with the high-end electrical receiver (8.8mV fixed noise), the critical length is found to be about 43cm with low optical coupling losses and close to ideal modulator at the bit rate of 6Gb/s. At higher bit rates and lower BER, the critical length reduces and optics becomes more power favorable. These trends can be fundamentally thought of in terms of a two-fold trade-off between electrical and optical interconnects. Optical interconnects are superior because they have lower attenuation and lower noise (no crosstalk etc). Whereas, their downside is that they need extra power for conversion from electronics to optics and vice versa. Since the power penalty is fixed, whereas, the power advantage is length and bit rate-dependent, the optical interconnects become beneficial at longer lengths. Finally, including several factors, which were ignored in this electrical interconnect analysis, such as equalization power, especially, as more taps are required at larger lengths and bit rates, incomplete residual ISI cancellation even with more taps, and rise time reduction induced greater package ringing at higher bit rates, will further reduce the critical lengths.

We have also quantified the impact of technology scaling on the critical length for bandwidth-sensitive applications who's interconnect lengths are in 10cm to 1m

range (chip-to-chip and board-level applications). We find that the transistor performance does not have a significant impact on the critical length, where as, the rapidly increasing bit rate dramatically lowers it. For optics to be favorable for short distances coupling loss and detector capacitance must be lowered for both lower laser power requirement and lower power dissipation. The critical length for 45nm technology node with 25fF detector capacitance is in the 10 to 20cm range, making it suitable for inter-chip communication in addition to longer backplane applications.

2.7 References

[1] R. T. Chen, L. Lin, C. Choi, Y. J. Liu, B. Bihari, L. Wu, S. Tang, R. Wickman, B. Picor, M. K. Hibbs-brenner, J. Bristow, and Y. S. Liu, "Fully embedded board-level guided-wave optoelectronic interconnects," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 780-793, June 2000.

[2] Y. Li, J. Ai, and J. Popelek, "Board-level 2-D data-capable optical interconnect circuits using polymer fiber-image guides," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 794-805, 2000.

[3] T. May, A. G. Kirk, D. V. Plant, J. F Ahadian, C. G. Fonstad, K. L. Lear, K. Tatah, M. S. Robinson, and J. A. Trezza, "Interconnection of a two-dimensional array of vertical-cavity surface-emitting lasers to a receiver array by means of a fiber image guide," *Applied Optics*, vol. 39, no. 5, pp. 683-689, Feb. 2000.

[4] K. P. Jackson, "High-density, array, optical interconnects for multi-chip modules," *Multi-Chip Module Conference*, pp. 142-145, Mar. 1992.

[5] G. A. Keeler, B. E. Nelson, D. Agarwal, and D. A. B. Miller, "Optical interconnects using short optical pulses," *Lasers and Electro-Optics Society 12th Annual Meeting*, 1999.

[6] Y. Liu, E. M. Strzelecka, J. Nohava, M. K. Hibbs-brenner, and E. Towe, "Smart-pixel array technology for free-space optical interconnects," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 764-768, 2000.

[7] D. V. Plant, and A. G. Kirk, "Optical interconnects at the chip and board level: challenges and solutions," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 806-818, 2000.

[8] S. C. Esener, "Implementation and prospects for chip-to-chip free-space optical interconnects," *International Electron Devices Meeting*, 2001.

[9] A. V. Mule, A. Naeemi, E. N. Glytsis, T. K. Gaylord, and J. D. Meindl, “Towards a comparison between chip-level optical interconnection and board-level exterconnection,” *International Interconnect Technology Conference*, pp. 92-94, June 2002.

[10] A. Naeemi, A. V. Mule, and J. D. Meindl, “Partition length between board-level electrical and optical interconnects,” *International Interconnect Technology Conference*, pp. 230-232, June 2003.

[11] D. A. B. Miller, “Rationale and challenges for optical interconnects to electronic chips,” *Proceedings of the IEEE*, vol. 88, no. 6, pp.728-749, 2000

[12] M. R. Feldman, S. C. Esener, C. C. Guest, and S. H. Lee, “Comparison between optical and electrical interconnects based on power and speed considerations,” *Applied Opt.*, vol. 27, no. 9, pp. 1742-1751, May 1988.

[13] M. Yoneyama, K. Takahata, T. Otsuji, and Y. Akazawa, “Analysis and application of a novel model for estimating power dissipation of optical interconnections as a function of transmission bit error rate,” *Journal of Lightwave Technology*, vol. 14, no. 1, pp. 13-22, 1996.

[14] O. Kibar, D. A. A. Blerkon, C. Fan, and S. C. Esener, “Power minimization and technology comparison for digital free-space optoelectronic interconnects,” *Journal of Lightwave Technology*, vol. 17, no. 4, pp. 546-555, April 1999.

[15] G. A. Keeler, N. C. Helman, P. Atanackovic, and D. A. B. Miller, “Cavity resonance tunneling of asymmetric Fabry-Perot MQW modulators following flip-chip bonding to silicon CMOS,” *Optics in Computing*, 2002.

[16] P. K. Tien, “Light waves in thin films and integrated optics,” *Applied optics*, vol. 10, no. 11, pp. 2395-2413, 1971.

[17] P. Kapur, and K. C. Saraswat, “Minimizing power dissipation in chip to chip optical interconnects using optimal modulators and laser power**,”** *International Interconnect Technology Conference*, pp. 224-226, June 2003.

[18] A. V. Krishnamoorthy, D. A. B. Miller, "Scaling optoelectronic-VLSI circuits into the 21st century: a technology roadmap," *Journal of Selected Topics in Quantum Electronics*, Vol. 2. No. 1, pp. 55-76, Apr. 1996.

[19] P. Kapur, and K. C. Saraswat, **"**Power dissipation in optical clock distribution network for high performance ICs," *International Interconnect Technology Conference*, pp. 151-153, 2002.

[20] *International Technology Roadmap for Semiconductor*, 1999ed, San Jose, CA: Semiconductor Industry Association.

[21] W. J. Dally and J. H. Poulton, *Digital systems engineering*, Cambridge University Press, NY, 1998.

[22] W. J. Dally, and J. Poulton, "Transmitter equalization for 4Gbps Signaling," *IEEE Micro*, pp. 48-56, 1997.

[23] K. Y. K. Chang, J. Wei, C. Huang, S. Li, K. Donnelly, M. Horowitz, Y. Li, S. Sidiropoulos, "A 0.4-4Gb/s CMOS quad transceiver cell using on-chip regulated dual-loop PLLs," *Journal of Solid-State Circuits*, vol. 38, no. 5, pp. 747-754, May 2003.

[24] B. K. Casper, H. Haycock, and R. Mooney, "An accurate and efficient method for multi-Gb/s chip-to-chip signaling schemes," *Symposium of VLSI circuits*, pp. 54-57, 2002.

[25] M. J. E. Lee, W. Dally, and P. Chiang, "A 90mW 4Gb/s equalized I/O circuits with input offset cancellation," *International Solid-State Circuits Conference*, pp. 252-253, 2000.

[26] C. Svensson, and G. D. Dermer, "Time domain modeling of lossy interconnects," *Transactions on Advanced Packaging*, vol. 23, no. 2, pp. 191-196, May 2001.

[27] D. A. B. Miller, and H. M. Ozaktas, "Limit to the bit-rate capacity of electrical interconnect from the aspect ratio of the system architecture," *Journal of Parallel and Distributed Computing*, vol. 41, 1997.

Chapter 3

Transmitter Technology: Quantum Well Modulator and VCSEL

3.1 Introduction

As we discussed in Chapter 2, optical interconnects can alleviate the impending off-chip electrical interconnect bottleneck by providing extremely dense, high-bandwidth links. They also have a potential of a lower power dissipation compared to their metal counterparts for inter-chip communications, especially as the technology (device dimension) scales down and the required bandwidth increases [1][2]. In Chapter 2 we assumed the QWM as a transmitter for optical links, however, an additional competing transmitter, the VCSEL, is available. There are many tradeoffs between these choices in terms of the performance, the integration complexity, and the reliability [3]. VCSEL simplify packaging by allowing the optical power to be generated on-chip. VCSEL can also provide a larger contrast ratio than modulators. The disadvantage stems from a poor reliability, especially in the high

temperature environment of a modern integrated circuit. Thermal variation across an array of VCSEL can cause threshold shifts which cause variation in optical power. Thermal crosstalk has been a huge issue in monolithic arrays of VCSEL, and it could be worse in flip-chip bonded arrays because the thermal conductivity between devices should be greater [4]. The VCSEL's bandwidth is limited by the ratio of operating current to the threshold current given by [5]

$$f_{3dB} = 1.55 f_o \left(\frac{I_{op}}{I_{th}} - 1 \right)^{1/2} \qquad (3.1)$$
$$f_o = 10GHz/\sqrt{mA} \text{ for } 850\text{nm devices}$$

where, I_{op} is the operating current and I_{th} is the threshold current of the VCSEL.

The maximum bandwidth also limited by thermal reliability aforementioned. QWMs, on the other hand, can be more reliable, can possibly be integrated monolithically in silicon, and do not dissipate large power, at least in the transmitter. QWMs' bandwidth is limited not by device itself, but parasitic bump capacitance, and QWMs have the higher bandwidth. QWMs can be programmed to be either detectors or modulators, which allows bidirectional data flow on a single channel [6]. However, their low contrast ratio, in particular, at the scaled CMOS supply voltages renders a higher power dissipation at the receiver-end. Although VCSEL have exhibited small-signal modulation bandwidths up to 21.5GHz, direct modulation may not provide adequate performance at data rate of 40Gb/s and beyond [7]. The monolithic integration of VCSEL and QWMs offers an attractive means of achieving higher modulation bandwidth. Recently, electroabsorption modulated laser (EML) consisting of a continuous wave (CW) laser and an electroabsorption modulator demonstrates operating at 40Gb/s [8][9].

In the past, there have been several attempts at quantifying the power dissipation of optical links. However, many researchers use fixed optoelectronic devices and receivers without considering their optimization [10][11]. Other researchers such as Kibar et. al. have optimized the power for the QWM and the VCSEL-based interconnects [12]. However, their comparisons assume fixed interconnect length and detector capacitance, and are limited only to the optical interconnects.

In this chapter, we use a different optimization scheme than [12], where we exploit the opposite power dissipation trends of the receiver and the transmitter with increasing laser power. In addition, we quantify the optimized power as a function of the link-length and establish bandwidth regimes where one technology is favorable over the other. Further, we compare both optical transmitter technologies with the state-of-the-art electrical interconnects whose noise sources and attenuation were modeled using simulations and numerical approaches. We define the critical length as the link length beyond which the optical interconnect is more power efficient than its electrical counterpart. Finally, we find the detector capacitance to be a critical parameter in determining the power dissipation of the optical links, hence quantify aforementioned comparison trends as a function of this parameter.

3.2 Power Modeling

Fig.3.1 shows the circuit schematic of the QWM and the VCSEL transmitters driven by a CMOS buffer chain. The buffer chain is assumed to be fan-out-of four to minimize the transmitter delay [12]. The power dissipation due to absorption of the

QWM is described in section 2.2.1. On the other hand, the electrical power dissipation in the VCSEL is the average static power (Fig.3.1b) given by

$$P_{VCSEL} = (i_{th} + i_m / 2)(V_{th} + R_s i_m + V_{swing} - V_{tn}) \tag{3.2}$$

where, R_s is the series resistance, i_{th} is the threshold current, i_m is the modulation current, V_{tn} is the NMOS threshold voltage. For a given laser slope efficiency (η_{LI}), the output optical power difference between the "on" and the "off" bits is

$$P_{O,VCSEL} = i_m \eta_{LI} \tag{3.3}$$

The dynamic power of the buffer chain is also included for both the VCSEL and the QWM.

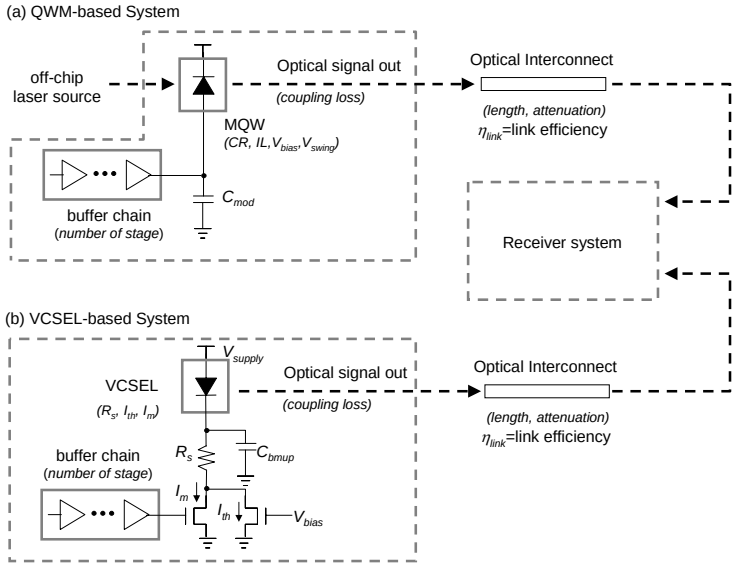


Fig.3.1 Schematics for (a) QWM and (b) VCSEL transmitters

For the receiver, we use a photo-detector followed by the transimpedance amplifier with subsequent gain stages. Its power modeling as well as the methodology to minimize the total optical link power has been detailed in section 2.2. The coupling and the medium losses were chosen to be 3dB and 0.082dB/cm, respectively [13].

Fig.3.2 shows the transmitter power dissipation comparison between the QWM and the VCSEL as a function of the required optical power difference at the receiver. The transmitter technology parameters for 850nm wavelength are shown in Table 3.1 [12]. The VCSEL has a standby power component due to threshold current (i_{th}), but, has a higher efficiency than the QWM for the chosen parameters. For the required optical power difference less than 110μW, the QWM dissipates less power. The required optical power difference at the receiver increases with a higher bandwidth and link length.

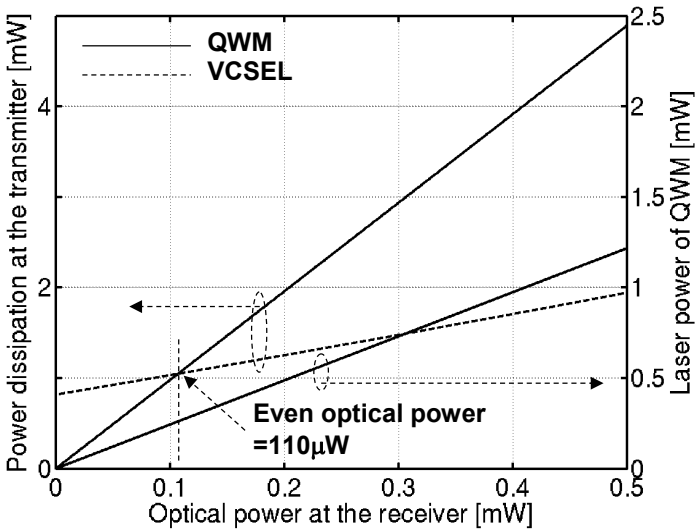


Fig.3.2 Transmitter power comparison in terms of the optical power at the receiver

Table 3.1 Transmitter technology parameters

QWM	*IL*	*CR*	V_{bias}	V_{swing}
	0.475	4.6	4.7 (V)	1.1 (V)
VCSEL	I_{th}	R_s	V_{th}	η_{LI}
	0.290 (mA)	250 (Ω)	2.0 (V)	0.7

3.3 Comparison between VCSEL and the QWM

The performance of the transmitter is a strong function of the bandwidth and the detector capacitance rather than the transistor performance because a small detector capacitance requires a lower optical power difference at the receiver to achieve the same signal to noise ratio [1][2]. In this work, we assume transistor performance corresponding to the 65nm technology node [14] with the bandwidth range from 5Gb/s to 30Gb/s, corresponding to the required bandwidth at the 90nm and the 45nm technology nodes, respectively. The detector capacitance is varied from 5fF (Metal-Semiconductor-Metal photo-detector) to 50fF (Flip-chip bonding technology [1]).

Fig.3.3 shows the optimized optical link power comparison as a function of the interconnect length for different detector capacitances at 10Gb/s. The even length, defined as the length beyond which VCSEL-based optical links are more power efficient than the QWM-based links, increases for lower detector capacitance, making the QWM more favorable. For detector capacitance more than 25fF, however, the VCSEL is more power efficient for the entire range of the interconnect lengths.

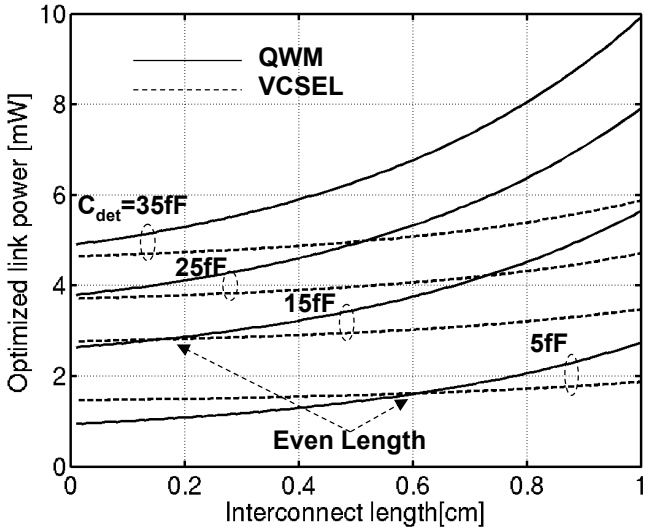


Fig.3.3 Optimized link power in terms of the interconnect length for different detector capacitances at 10Gb/s

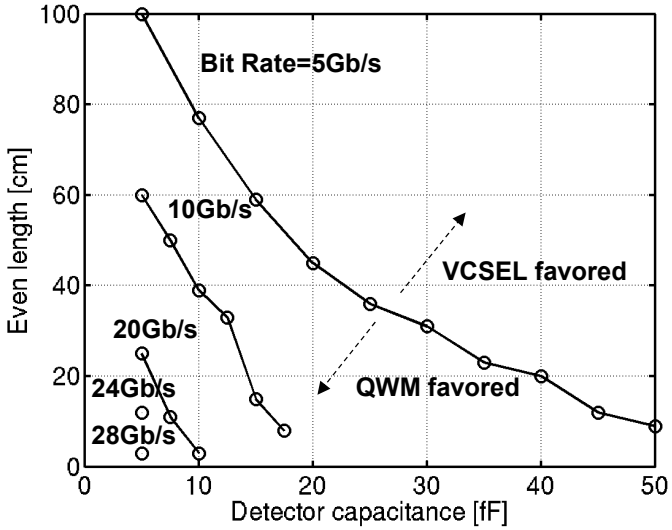


Fig.3.4 Even length in terms of the detector capacitance for different bandwidths

Fig.3.4 shows the even length in terms of the detector capacitance and bit rate. For moderate detector capacitances (25fF) and low bandwidths (5Gb/s), the even

length is more than 35cm, indicating that the QWM would be favorable for short-distance interconnects. However, for very high bandwidths, this length reduces to less than 10cm.

3.4 Comparison with Electrical Counterpart and Power Gain

In the Chapter 2, we have analyzed the impacts of system and device parameters on the power dissipation for the electrical and only the QWM-based optical interconnects. We extend this comparison to include the impact of different optical transmitter technologies. For the electrical interconnect, simultaneous bi-directional signaling scheme with a very low-loss tangent dielectric material (GETEK board) is assumed [1] as described in the section 2.3.

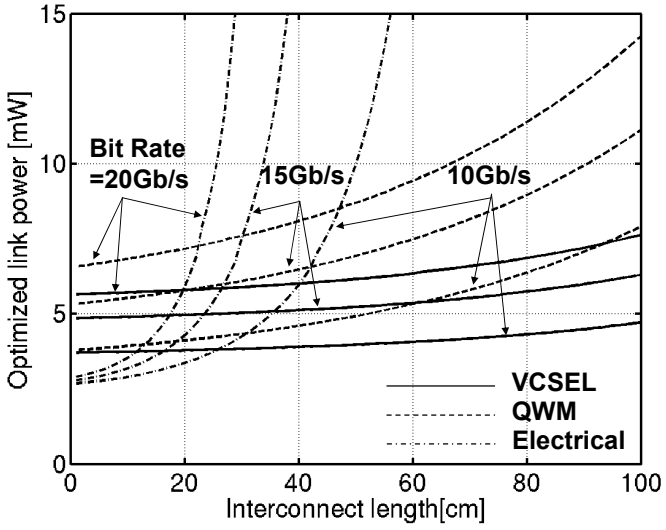


Fig.3.5 Power comparison between the electrical and the optical interconnect showing the critical lengths at a detector capacitance of 25fF

Fig.3.5 shows the link power dissipation as a function of its length for different bandwidths. Beyond the critical length, optical interconnects yield lower power dissipation. This length ranges from 20~30cm for both the QWM and the VCSEL technologies at 10~20Gb/s with a 25fF detector capacitance.

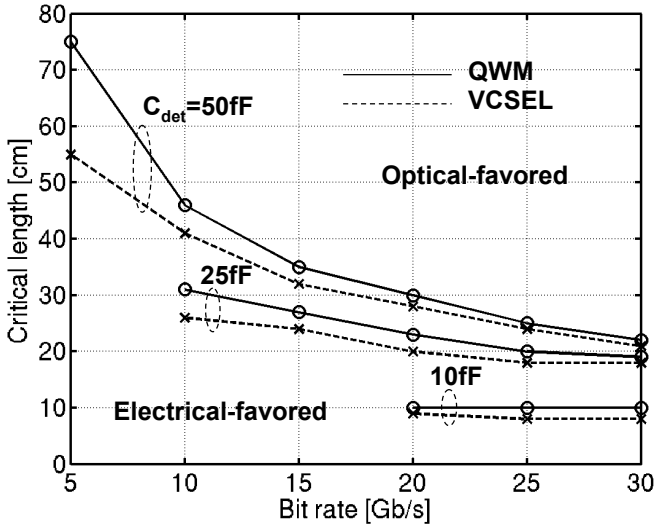


Fig.3.6 Critical length in terms of system bit rate for the QWM and the VCSEL transmitter technologies

Fig.3.6 summarizes the impact of the detector capacitance and the bandwidth on the critical length with different transmitter technologies. The QWM critical length is slightly larger. The difference is larger at a lower bit rate and a higher detector capacitance. For high bit rates and low detector capacitances, the critical length is similar as the electrical power rises dramatically with the length at high bit rates. This indicates both the QWM and the VCSEL transmitter technologies can out-perform the electrical interconnect for inter-chip communications at the future nodes, especially at a high required system bandwidth and if low detector capacitances are used. Fig.3.7 shows the power gain of the VCSEL compared to the QWM at the critical length. For 50fF of detector capacitance, the gain with the VCSEL is 35% at 5Gb/s and saturates

to 23% after 15Gb/s. For 25fF of detector capacitance, the gain increases less than 20Gb/s, subsequently, saturating as the previous case. 10fF exhibits a similar trend as the 25fF, but the QWM gets close to the VCSEL before 20Gb/s.

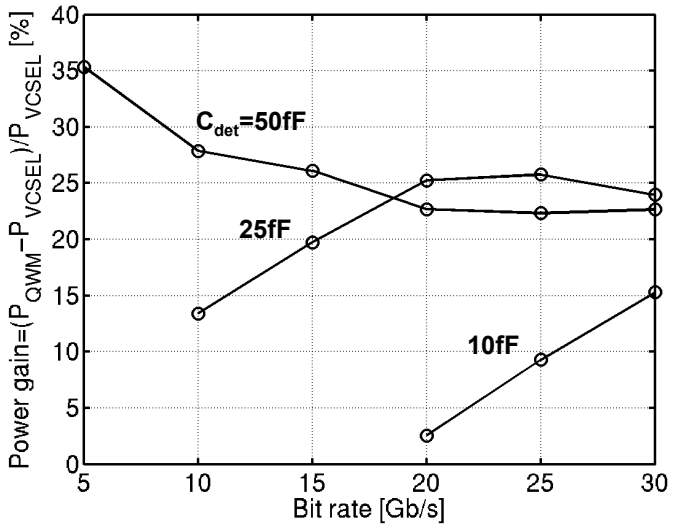


Fig.3.7 Power gain of the VCSEL compared to the QWM

3.5 Modulator Trade-off: IL, CR, and Laser Power

As we discuss in section 3.3 and 3.4, QWM is the transmitter technology of choice at lower capacitance and bandwidth, whereas VCSEL is more power efficient at higher bandwidth. However, high modulation bandwidth of VCSEL may not provide adequate performance at data rate of 40Gb/s and beyond [7]. Hence, it is worthwhile examining the trade-off among metrics of the modulator (such as insertion loss, contrast ratio, and laser power) to extend the range of capacitance and bandwidth

over the VCSEL. In this chapter, we further examine the impact of these modulator metrics on the comparison.

3.5.1 Evaluating Total Link Power Dissipation with MQW's Design Parameters

Fig.3.1 shows the schematic of QWM and VCSEL transmitters driven by the delay-optimized fan-out four CMOS buffer chains. The MQW power dissipation due to absorption is given by (2.1). The QWM metrics, which critically impact the link performance, are its capacitance, *IL*, *CR*, and the ratio between the maximum and minimum absorption (*X*). For the Fabry-Perot cavity modulator, the *IL* and *CR* tradeoff is given by

$$CR = \left(r_{on} \frac{1 + X(1 - r_{on})/(1 + r_{on})}{1 - X(1 - r_{on})/(1 + r_{on})} \right)^2 \quad , \; r_{on} = \sqrt{1 - IL} \tag{3.4}$$

where r_{on} represents the field reflectivity for "on" state [2]. The electrical power dissipation in the VCSEL is the average static power given by (3.2).

Fig.3.8 shows the transmitter power dissipation comparison as a function of required optical power difference at the receiver at a given QWM parameters (V_{bias}=3V, *X*=5, and different *IL*) and 10Gb/s of bit rate. VCSEL has a standby power component due to threshold current (i_{th}), but has a higher efficiency than QWM for the chosen parameters. The crossover optical power difference at the receiver, henceforth referred to even optical power, increases for small IL as shown in Fig.3.9. However, the receiver power has opposite trends in terms of optical power difference at the

receiver hence the optimized optical power determined by a link length and receiver metrics to minimize the total link power [1]. We evaluate the optimized receiver optical power difference for both transmitter technologies and establish bandwidth regimes where one technology is favorable over the other with modulator's design constraints.

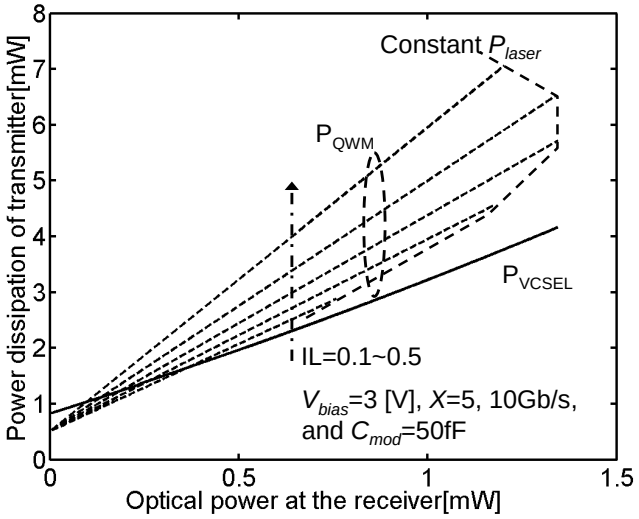


Fig.3.8 Transmitter power dissipation at a given QWM metrics

3.5.2 Comparison between VCSEL and QWM-based Optical Link

The performance of the transmitter is a strong function of the bandwidth and its capacitance because smaller transmitter capacitance reduces the dynamic power and smaller detector capacitance requires lower optical power at the receiver [1]. In this work, we assume transistor performance corresponding to 65nm technology node [14] with bandwidth range from 5Gb/s to 30Gb/s, corresponding to the required

bandwidth at 90nm and 45nm technology nodes, respectively. The capacitance is varied from 10fF (achievable with MSM photo-detector and monolithic QWM) to 50fF (Flip-chip bonding technology [15]).

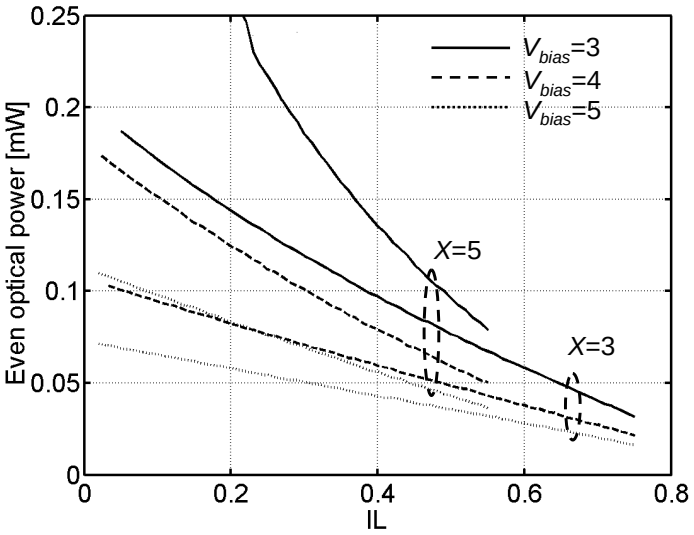


Fig.3.9 Even optical power as a function of QWM metrics at 10Gb/s

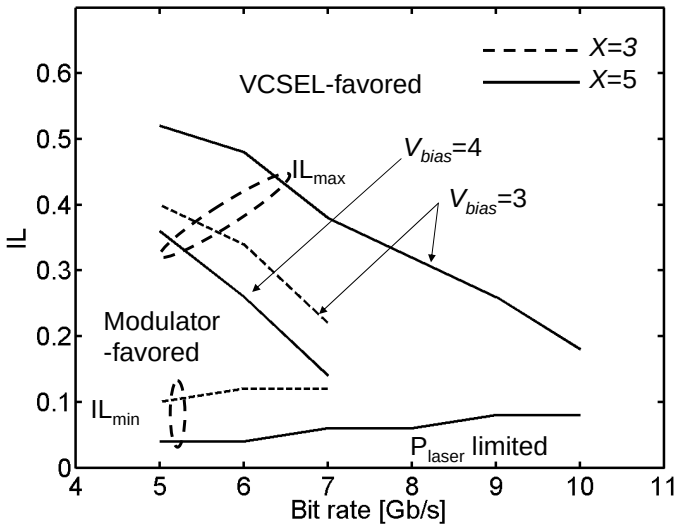


Fig.3.10 *IL* as a function of bit rate for link length=50cm (inter-board) and C_{mod}=C_{det} =50fF

Fig.3.10 shows the range of IL as a function of bit rate with different MQW design parameters for 50fF of both transmitter and receiver capacitances at the 50cm of link length (inter-board communication). The maximum IL is determined by where MQW-based links have same power dissipation with VCSEL-based links. On the other hand, lower IL requires larger input laser power (in this work, 1mW), hence it limits the minimum IL as shown in Fig.3.10. The MQW's achievable bandwidth regime is around 10Gb/s for X=5 and V_{bias}=3V, best case in our simulation range.

Fig.3.11 shows the same plot at the 10cm of link length (inter-chip communication). The lower loss increases can tolerate the higher IL to get the same optical power at the receiver. However, the bandwidth regime extends only couple of Gb/s because the dynamic power of the MQW for high capacitance (50fF) becomes dominant for higher bandwidth.

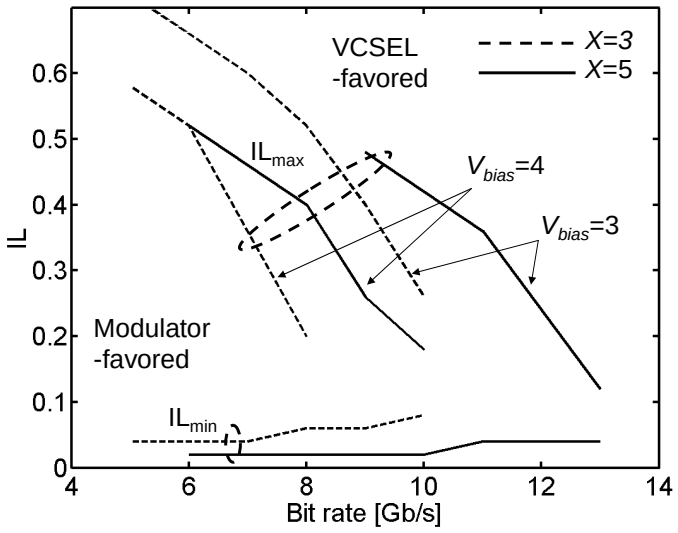


Fig.3.11 *IL* as a function of bit rate for link length=10cm (inter-chip) and C_{mod}=C_{det} =50fF

Fig.3.12 shows the impact of transmitter capacitance on these trends aforementioned for X=5 and V_{bias}=3V. For 10fF of transmitter capacitance, the bandwidth regime gains beyond 30Gb/s with reasonable range of IL. For VCSELs, dynamic power is only coming from driving NMOS transistor (providing I_m) capacitance without flip-chip bond pad capacitance, comparable to 10~15fF of MQW's capacitances. Fig.3.12 also shows the impact of detector capacitances, extending the bandwidth regime.

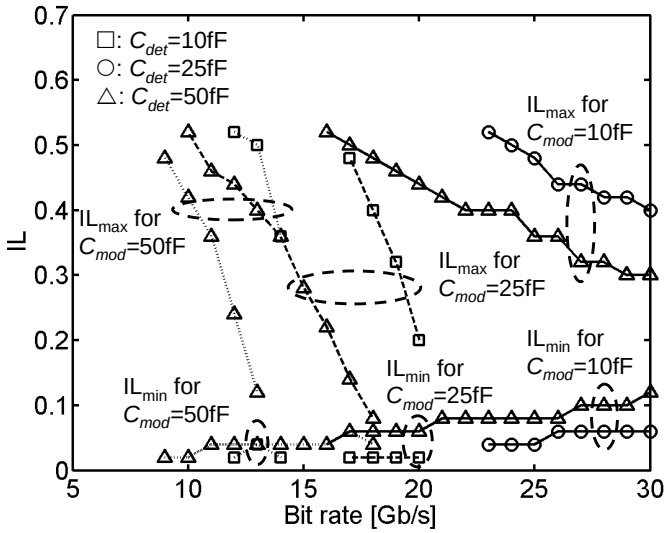


Fig.3.12 Impact of transmitter capacitance for X=5 and V_{bias}=3V.

3.6 Summary

In the first of this chapter, we quantify and compare the power dissipation of two competing transmitter technologies: VCSEL and QWM for the inter-chip communication using high speed optical links. VCSEL-based links are power

favorable compared to QWM-based links at higher bandwidth (>20Gb/s) and larger distance. However, the QWM is better for lower detector and modulator capacitances. We also extend the comparison to high-speed electrical links and quantify the critical length beyond which optical links are power favorable compared electrical links for both VCSEL and QWM transmitters.

In the second part, we quantify the design constraints on the modulator under which it is superior to VCSEL technology as a function of bandwidth, link length, and transmitter and detector capacitances. We also identify the range of tolerable insertion loss in QWM. To render QWM more power-efficient, lower capacitances (<15fF) of both the modulator and the photo-detector would be demanded. In the next chapter, we discuss the optimization of the modulator using semi-empirical model of the modulator absorption characteristic.

3.7 Reference

[1] Hoyeol Cho, Pawan Kapur, and Krishna C. Saraswat, "Power Comparison between High-speed Electrical and Optical Interconnects for Inter-chip Communication," *Journal of Lightwave Technology*, vol.22, no.9, pp.2021-2033, Sept. 2004.

[2] Hoyeol Cho, Pawan Kapur, and Krishna C. Saraswat, "Impact of Technology Node on Power Comparison for High-Speed Electrical and Optical Interconnects," *IEEE International Interconnect Technology Conference*, pp.177-179, June 2005.

[3] A. L. Lentine, "Qualitative comparison of MQW modulator and VCSEL based OE-VLSI: a systems perspective," *LEOS 10th Annual Meeting*, vol. 2, pp. 83-84, Nov. 1997.

[4] T. Wipiejewski, D. B. Young, B. J. Thibeault, and L. A. Coldren, "Thermal crosstalk in 4×4 vertical-cavity surface-emitting laser arrays," *IEEE Photonic Technology Letters*, vol. 8, no.8, pp. 980-982, Aug. 1996.

[5] K. Lear, Chip to Chip Optical Interconnect Kickoff Meeting, June, 2003.

[6] A. V. Krishnamoorthy, T. K. Woodward, K. W. Goossen, J. A. Walker, S. P. Hui, B. Tseng, J. E. Cunningham, W. Y. Jan, F. E. Kiamilev, and D. A. B. Miller, "Dual-function detector-modulator smart-pixel module," *Applied Optics*, vol. 36, pp. 4866-4870, 1997.

[7] D. K. Serkland, G. M. Peake, and K. M. Geib, "VCSEL modulation using an integrated electro-absorption modulator," *Laser and Electro-Optics conference on*, vol. 1, pp. 16-21, May 2004.

[8] R. Nagarajan et al., "Large-scale photonic integrated circuits," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 11, no. 1, pp. 50-65, Jan. 2005.

[9] [On-line] Infinera News available at http://www.infinera.com/news/2006-03-27.html.

[10] A. V. Krishnamoorthy and D. A. B. Miller, "Scaling optoelectronic VLSI circuits into the 21st century: A technology roadmap," *Proceeding of Journal of Selected Topics Quantum Electronics*, vol.2, no.1, pp.55-67, Apr. 1996.

[11] M. Yoneyama, K. Takahata, T. Otsuji, and Y. Akazawa, "Analysis and application of a novel model for estimating power dissipation of optical interconnects as function of transmission bit error rate," *Journal of Lightwave Technology*, vol.14, no.1, pp.13-22, Jan. 1996.

[12] Osman Kibar, Daniel A. Van Blerkom, Chi Fan, and Sadik C. Esener, "Power Minimization and Technology Comparisons for Digital Free-Space Optoelectronic Interconnects," *Journal of Lightwave Technology*, vol. 17, no.4, pp. 546-554, Apr. 1999.

[13] P. K. Tien, "Light waves in thin films and integrated optics," *Applied Optics*, vol. 10, no. 11, pp. 2395–2413, 1971.

[14] "The International Technology Roadmap for Semiconductor (2004 Edition),"

[15] G. A. Keeler, N. C. Helman, P. Atanackovic, and D.A.B. Miller, "Cavity resonance tunneling of asymmetric Fabry-Perot MQW modulators following flip-chip bonding to silicon CMOS," *Optics in Computing*, 2002.

Chapter 4

QWM Design Methodology

4.1 Introduction

The communication bandwidth demand for short distance interconnects including applications such as inter-chip or inter-board interconnects is rising dramatically. The International Technology Roadmap for Semiconductors (ITRS) predicts that the bandwidth demand will increase much faster than even the improvement in the transistor performance [1]. Traditional copper (Cu) board traces are unable to keep up with the steep demand curve due to deteriorating signal integrity at the receiver arising from rapid signal attenuation at high frequencies (skin effect and dielectric loss), increase in cross-talk, and reflections from impedance mismatch and package parasitics [2]-[4]. The mitigating solutions such as low loss-tangent dielectric materials and better signal equalization schemes [5]-[7] are both expensive as well as require large power and area. Optical interconnects present an attractive alternative, providing high bandwidth potentially at lower power [8]. However, for short distance high-speed optical links to be competitive enough to replace electrical links, design

methodologies minimizing their power dissipation must be developed. This includes optimization with respect to device (number of quantum wells, capacitances) as well as system parameters (bit rate, pre-bias and swing voltages, link efficiency, and input laser power).

In the past, a methodology maximizing a figure of merit given by the product of the reflectivity difference of the modulator and the responsivity of the detector was developed by Neilson for quantum confined stark effect based devices [9]. However, it did not directly address power dissipation. Its focus was on optimizing optoelectronic devices, hence, the impact of post-detector amplifying stage and modulator driving stage was not considered. Since these electronic stages dissipate a large fraction of the total power, they should be included in power optimization. Another research group optimized both modulator and VCSEL based links including the post-detector receiver stages [10] However, they assumed fixed modulator and detector properties and did not optimize the modulator. Subsequently, Kapur et. al. optimized modulator parameters (insertion loss and contrast ratio) and presented the feasibility of CMOS compatible low-voltage operation with modulator [11]. However, this optimization with its fixed modulator pre-bias condition and fixed voltage swing was applicable to low bit-rates (2-7 Gb/s), where receiver power was dominant. At high bit rates, the dynamic power at the transmitter becomes important, hence, modulator parameters (insertion loss and contrast ratio) characterized by pre-bias and swing voltages should be included in the optimization.

In this chapter, we propose an alternate, comprehensive, full link-aware power minimization framework (Fig.4.1), which addresses the aforementioned shortcomings. We use the GaAs-AlGaAs quantum well material system as an example. The methodology is widely applicable to generic quantum well material systems. A semi-empirical model for field dependence of the absorption edge is used to extract the

modulator properties in terms of pre-bias and swing voltage [9]. We optimize the modulator design parameters (swing voltage, pre-bias, and number of quantum wells), while considering the impact of receiver amplifier properties such as noise, bandwidth, and gain stages. We also present the feasibility and the power penalty associated with sub-optimum, low-voltage operation. Finally, we also quantify the impact of design parameters (modulator/detector capacitance) and the system parameters (bit rate, link efficiency, and off-chip laser power) on the optimization. The organization of the chapter is as follows. In section 4.2, we discuss the modeling of the quantum well modulator, using semi-empirical field dependence of the absorption edge. In section 4.4, we discuss the power optimization methodology and the corresponding optimum modulator design and operation parameters. In sections 4.4 and 4.5, we quantify the impact of system parameters and technology scaling on optimization, respectively. In section 4.6, we apply our optimization methodology to electroabsorption modulated lasers (EML). Finally we summarize in section 4.7.

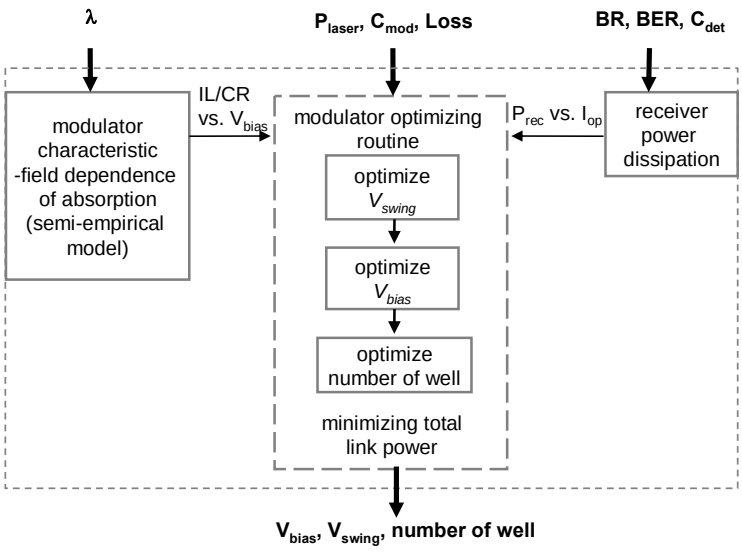


Fig.4.1 Power optimization sequence including modulator characteristic and receiver power dissipation routine providing device parameters

4.2 Semi-empirical Absorption Model

The modulator metrics, which impact the link performance and power, are its capacitance (C_{mod}), *IL*, and *CR*. The capacitance affects the modulator's dynamic power dissipation, while *IL* and *CR* impact both the receiver and the modulator power dissipation, and are related to the physical design of the modulator. The modulator device structure considered in this work is the p-i-n diode. The intrinsic region formed with N period of quantum wells acts as the absorption region. The absorption is modulated by an applied electric field (*F*) through quantum confined stark effect (Fig.4.2).

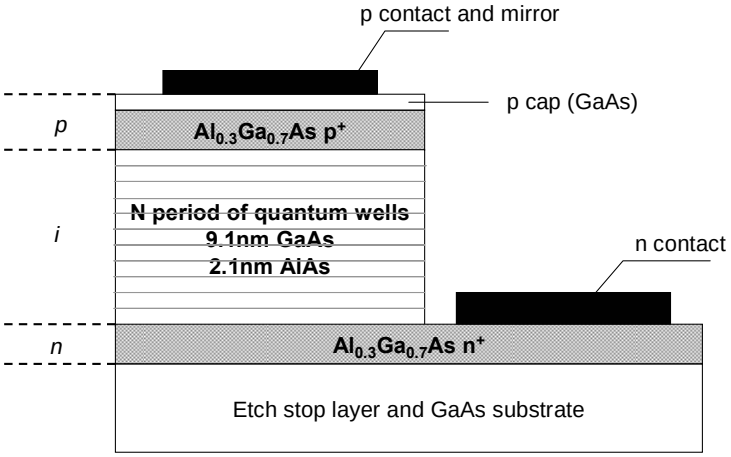


Fig.4.2 Modulator device structure (p-i-n diode) considered in this work

The dependence of absorption, α, on the electric field has been shown to be reasonably fitted using hyperbolic secant function in GaAs-AlGaAs QW's [9]. The α value at a given photon energy ($h\omega$) due to exciton, can thus be modeled as

$$\alpha = \alpha_p \text{sech}\left(1.317\frac{h\omega - E_{ex}}{\beta}\right) \tag{4.1}$$

where β is the half maximum of the exciton, E_{ex} is the field dependent energy of the exciton, and α_p is the peak absorption. These parameters have been shown to be fitted with linear and parabolic equations.

$$\begin{aligned} E_{ex} &= E_0 - E_1 F - E_2 F^2 \\ \alpha_p &= 1/\beta(\alpha_0 - \alpha_1 F - \alpha_2 F^2) \\ \beta &= \beta_0 + \beta_1 F + \beta_2 F^2 \end{aligned} \tag{4.2}$$

As an example, Fig.4.3 shows the result of field dependent absorption model for a specific case of 60 periods 9.1nm GaAs well and 2.1nm AlAs barrier device. The fitting parameters shown in Table 4.1 were taken from [9] and were fitted to the experimental data from [12].

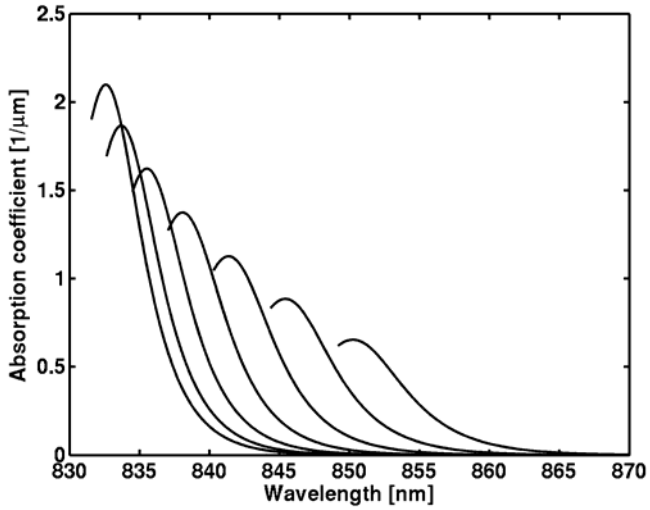


Fig.4.3 QW absorption coefficient of a 60-period GaAs-AlAs modulator with 9.1nm wells and 2.2nm barriers for applied voltages of 0-12V in 2V steps.

Table 4.1 Fitting parameters for field dependent absorption edge

Parameter	Value	Parameter	Value	Parameter	Value
E_0(meV)	1483.8	α_0(meVμm^{-1})	11.7	β_0(meVμm^{-1})	5.2
E_1(meVμmV^{-1})	0.15	α_1(meVV^{-1})	0.25	β_1(meVV^{-1})	0.05
E_2(meVμm^{2}V^2)	0.072	α_2(meVμmV^2)	0.005	β_2(meVμmV^2)	0.0027

We consider the reflective mode quantum well modulator. For this device, the *IL* and *CR* are given by

$$IL = 1 - R_{on} = 1 - e^{\alpha_{\min} 2l}$$
$$CR = \frac{R_{on}}{R_{off}} = \frac{e^{\alpha_{\min} 2l}}{e^{\alpha_{\max} 2l}} \tag{4.3}$$

Here, R_{on} and R_{off} are the modulator reflectivities in the less and the more absorbing states, respectively, and l is the total length of the absorbing well region. l corresponds to the thickness of the intrinsic (i) region associated with the number of wells. The absorption coefficient is assumed to have a linear dependency with the number of wells. [9] The power dissipation of the modulator due to absorption is evaluated with plugging equation (4.1) to equation (2.1). Dynamic power assuming 50% switching activity is given by

$$P_{dynamic,\mathrm{mod}} = 0.5(C_{\mathrm{mod}} + \sum_{i=1}^{nstage}(C_{ing}^{i} + C_{jo}^{i}))V_{swing}^{2} f \tag{4.5}$$

Here, C_{mod} includes the solder bump (flip-chip bonded modulator) and the modulator capacitance and C_{ing}^{i} and C_{jo}^{i} are the input and output capacitance of i^{th} stage, respectively. The buffer chain is assumed to be fan-out-of-four to minimize the transmitter delay [13]. The transistor parameters (65nm technology node) for receiver

noise modeling were chosen from the ITRS [1].

4.3 Power Optimization Methodology

Fig.4.4 and Fig.4.5 illustrate our optical power minimization methodology. Fig.4.4 shows the optimum voltage swing for a given pre-bias modulator voltage. Increasing the swing voltage reduces two out of the three power dissipation components, while increasing the third. Firstly, it reduces receiver power by increases the optical power difference between the "on" and the "off" states. Thus, allowing a larger noise for the same signal to noise ratio. Secondly, it decreases the static power of the modulator since insertion loss decreases due to low absorption in the "on" state. On the other hand, a higher voltage swing increases the dynamic power of the modulator in a quadratic manner. This leads to an optimal V_{swing} minimizing total interconnect power (receiver and modulator) at a given pre-bias voltage (V_{bias}) (Fig.4.4).

Fig.4.5 illustrates the existence of an optimum with respect to the modulator V_{bias}. Each V_{bias} in this figure has its respective optimum voltage swing as exemplified in Fig.4.4. Fig.4.5 exhibits three distinct regions in the total link power with respect to V_{bias}. For a small V_{bias} (region I), the dynamic and the static powers of the modulator are small whereas, a low *CR* at the receiver leads to a higher receiver power dissipation, which becomes dominant and results in an increase in the total link power. For intermediate V_{bias} (region II), the three powers (static/dynamic power of the modulator, receiver power) are comparable. However, the receiver power is relatively independent of V_{bias} in this range, hence, the total link power exhibits a flat trend with V_{bias}. For high V_{bias} (region III), the static and the dynamic modulator power becomes

dominant, hence increases the optimized link power. For 20Gb/s (Fig.4.5), the optimized V_{bias} has a wide range. However as the bit rate increases, the optimized range becomes narrower, since the dynamic power increases and pushes region III to lower V_{bias}.

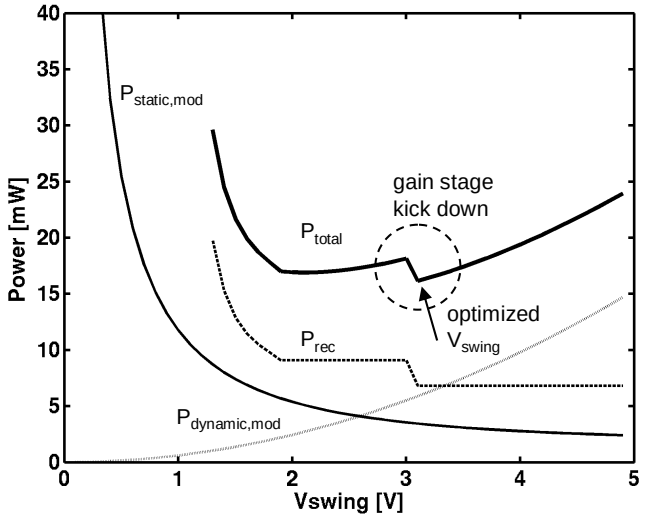


Fig.4.4 Link power vs. $V_{swing,}$ where, $P_{static,mod}$ is the static power of the modulator, $P_{dynamic,mod}$ is the dynamic power of the modulator, P_{rec} is the receiver power, and P_{total} is the total link power. An optimum V_{swing}, minimizing the total link power, is clearly observed. (C_{det}=C_{mod}=50fF, loss=3dB, laser power (P_{laser})=1mW, bit rate=20Gb/s, V_{bias}=5V, and 60 wells).

In this work, the operation wavelength is chosen to be 844nm for GaAs-AlGaAs QW's [9], which yields the lowest link power dissipation as shown in Fig.4.6. For wavelengths slightly shorter than 844nm, V_{bias} and V_{swing} increase to provide required reflectivity difference at the receiver, which results in increase of dynamic power at the modulator. For slightly longer wavelengths, the achievable reflectivity difference becomes small due to the broadening of the absorption as shown in Fig.4.3, which burns a large power at the receiver. However, the total link power is relatively

flat around 850nm (842nm~852nm). Fig.4.6 also indicates that this optimized wavelength is valid for different numbers of quantum wells.

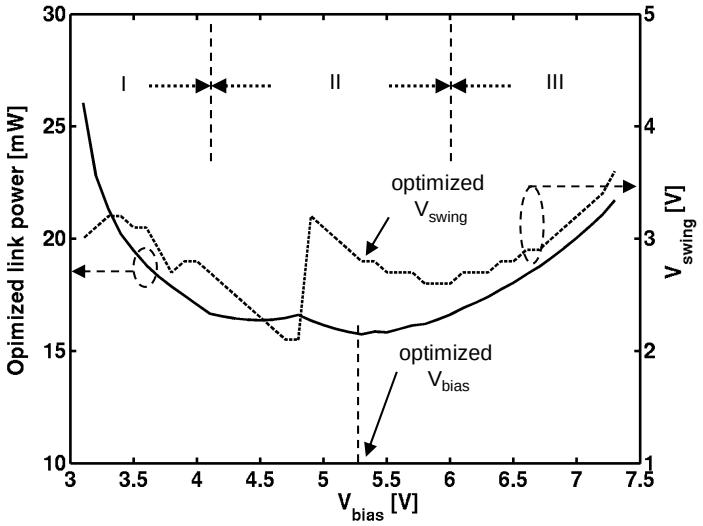


Fig.4.5 Optimum V_{swing} and the link power vs. V_{bias}. For every V_{bias}, the optimized V_{swing} is extracted along the lines of Fig.4.4. At V_{bias}=5V, the gain stage is reduced from four to three stages, increasing the optimum V_{swing}. The optimum V_{bias}, minimizing the total link power, is also seen. (C_{det}=C_{mod}=50fF, loss=3dB, laser power (P_{laser})=1mW, bit rate=20Gb/s, and 60 wells).

So far, we have found the optimum pre-bias and swing voltage for a given number of quantum wells in the intrinsic region of a modulator. Varying the intrinsic region thickness, thus, the number of quantum wells changes modulator's *IL*, *CR*), thus also affecting its optimum pre-bias and swing voltage. Fig.4.7 and Fig.4.8 show that there is a clear optimum in the total link power with respect to the total number of quantum wells as well. Each point on this graph is obtained using the preceding optimization with respect to V_{bias} and swing voltage. The optimum values for these parameters are also shown.

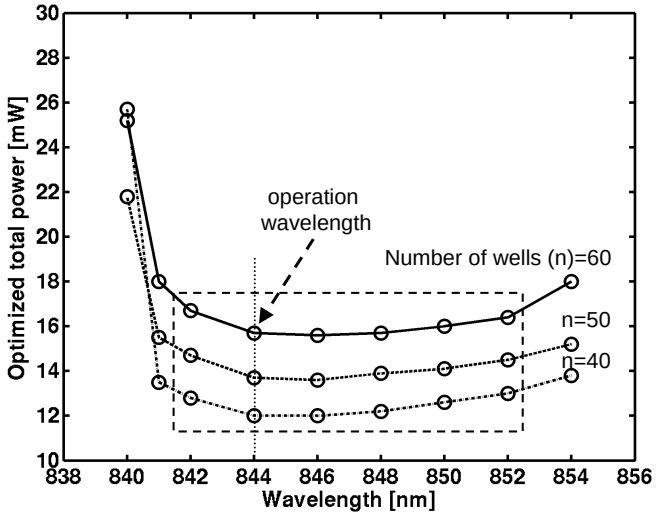


Fig.4.6 Optimized link power vs. wavelength. The figure shows an operating wavelength, which minimizes the total link power (C_{det}=C_{mod}=50fF, loss=3dB, laser power (P_{laser})=1mW, and bit rate=20Gb/s).

A small number of wells results in a low *CR* of the modulator. This, in turn, leads to a low "on"/"off" optical power difference at the receiver resulting in a large receiver power dissipation. Hence, the optimized link power increases as the number of wells decrease. On the other hand, a very large number of wells also results in a high total link power dissipation. In this case, the required pre-bias and the voltage swing to achieve reasonable electric field difference between "on" and "off" state increases. This results in a larger static and dynamic power in the modulator. Thus, there exists an optimum in total link power with respect to number of quantum wells. Comparison between Fig.4.7 and Fig.4.8 shows that a higher bit rate requires a larger number of quantum wells, and, as anticipated, burns a higher power at the optimum point.

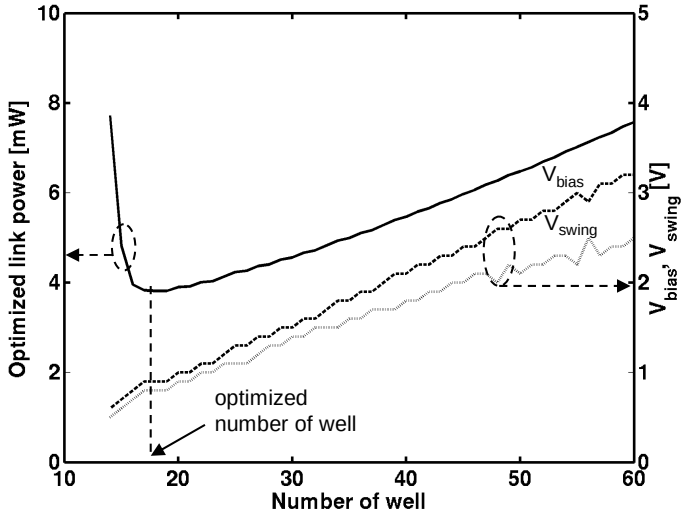


Fig.4.7 Optimized link power vs. number of quantum wells in the modulator. The figure clearly shows an optimum number of wells, V_{bias}, and V_{swing}, which minimizes the total link power (C_{det}=C_{mod}=50fF, loss=3dB, laser power (P_{laser})=1mW, and bit rate=10Gb/s).

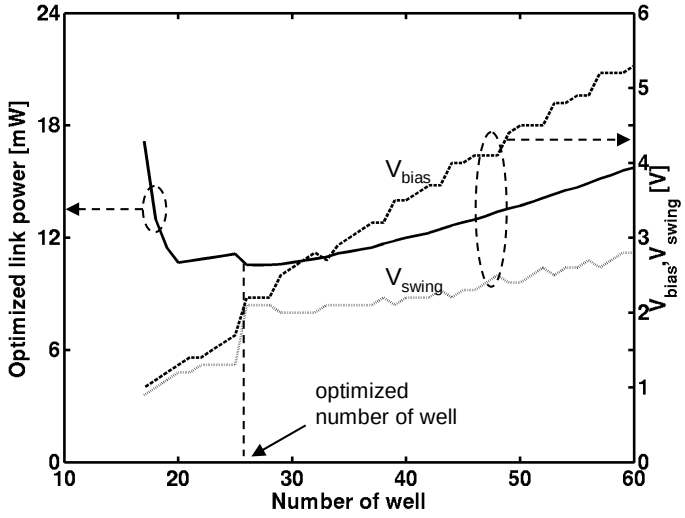


Fig.4.8 Optimum link power vs. number of wells of the modulator. The figure has same parameters as in Fig.4.7, except that the bit rate is now 20Gb/s

In order to be compatible with scaled operating voltages (V_{CMOS}=1.1V at 65nm technology node, ITRS [1]) of future CMOS or I/O compatible value (V_{IO}=1.8V), the modulators must operate at one of these voltages instead of the optimum values obtained above. This increases the link power, hence increases the power penalty. The power penalty is given by

$$\text{Power Penalty} = \frac{P(V_{swing} = \text{V}_{\text{CMOS}}, V_{IO}) - P(V_{swing,opt})}{P(V_{vswing,opt})} \times 100 \qquad (4.6)$$

where, $P(V_{swing}=V_{cmos},V_{IO})$ is the link power at V_{CMOS} and V_{IO}, and $P(V_{swing,opt})$ is the link power at optimum swing voltage. Fig.4.9 depicts the power penalty incurred when the voltage swing used is either a CMOS device compatible value at1.1V (65nm node, ITRS) or I/O compatible value at 1.8V instead of the optimum values obtained above [1]. For example, at 10Gb/s, the optimum swing voltage is 0.8V, which is lower than both the CMOS and the I/O supply voltages, hence power penalty incurred is approximately 8% and 34%, respectively. Optimum swing voltage increases with bit rate. For example, it is about 1.5V at 25Gb/s. Thus, there is a lower power penalty by operating at the I/O voltage instead of the CMOS voltage, opposite to the low bit rate case. For very high bit rates (>30Gb/s), the modulator could not be operated at voltage swing less than 1.8V, indicating a need for either a different power supply or internally pumping up the voltage for the modulator.

4.4 Impact of Device Parameters on Optimization

The optical communication system as shown in Fig.2.1 comprises several optical and electrical components representing device parameters, such as modulator

and detector capacitances, as well as system parameters, such as laser power, medium and coupling losses, and the bit rate. In this section, we examine the impact these parameters on the optimization (number of wells, pre-bias, and voltage swing).

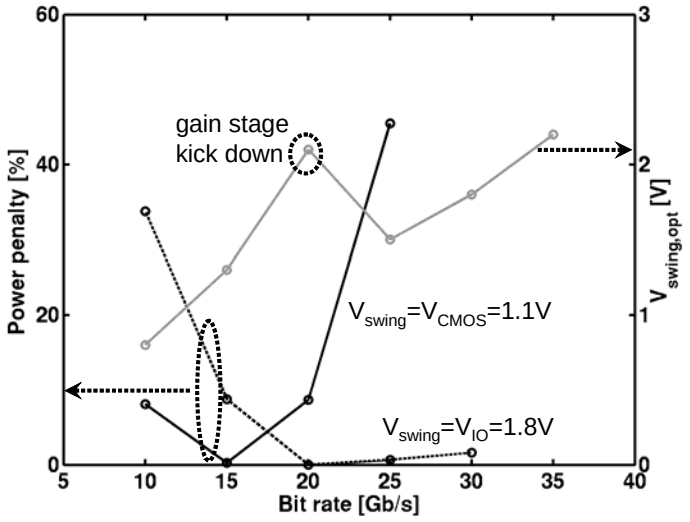


Fig.4.9 Power penalty of predicted CMOS supply (V_{CMOS}=1.1V) and I/O voltage (V_{IO}=1.8V) compared to optimize swing voltage ($V_{swing,opt}$) in terms of bit rate for C_{det}=C_{mod}=50fF, loss=3dB, and laser power (P_{laser})=1mW.

In Fig.4.10, we plot the optimum number of wells, pre-bias, and voltage swing as a function of bit rate. For higher bit rate, the required optical power at the receiver increases due to increase in noise components at the front-end of the receiver [14]. This requires a larger number of wells to provide enough *IL* and *CR*, which, in turn, demands a high pre-bias and swing voltage. For 20Gb/s, the receiver reduces the power with decreasing the gain stages since for this case, the receiver is in the gain-limited range, where the reduction in the power dissipation (reduction in the width) is no longer forbidden by a higher noise, but because of a low gain, hence, the minimum power per amplifier stage determined by the smallest possible transistor width, occurs

at an optical power difference corresponding to the beginning of the gain-limited region and above this optical power difference value power dissipation per stage remains constant, as shown in Fig.4.4.

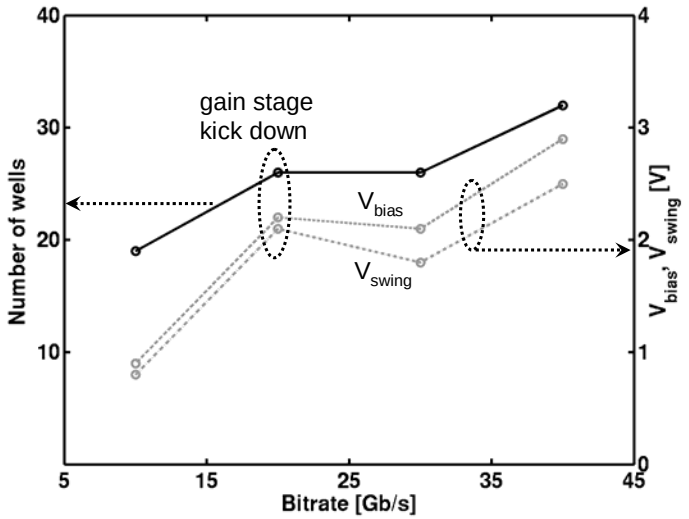


Fig.4.10 Impact of bit rate on the modulator design parameters: number of wells, V_{bias}, and V_{swing} (C_{det}=C_{mod}=50fF, loss=3dB, and laser power (P_{laser})=1mW)

Fig.4.11 plots the three optimum parameters at 20Gb/s as a function of the total optical power loss in the link from the transmitter to the receiver. Increasing loss has a similar effect as increasing bit rate since a higher loss reduces optical power at the receiver. The resulting optimal link power as a function of the loss and bit rate is shown in Fig.4.12. Bit rate, compared to the optical link efficiency, has a larger impact on the total link power dissipation. A higher bit rate increases both the transmitter and the receiver power. In contrast, a lower link efficiency (higher loss) only increases the transmitter power by requiring a larger optical power at the transmitter. A factor of 4 increase in bit rate (10 to 40Gbps) results in approximately an 8X increase in total link power.

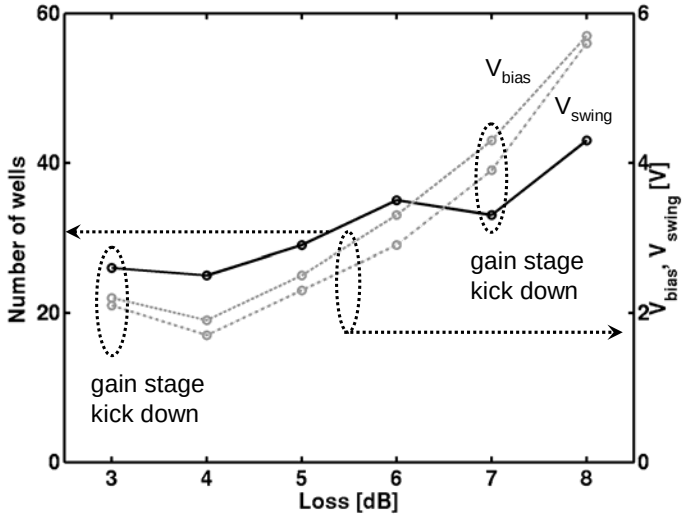


Fig.4.11 Impact of link efficiency (optical power loss) on the optimum modulator design parameters: number of wells, V_{bias}, and V_{swing} (C_{det}=C_{mod}=50fF, laser power (P_{laser})=1mW, and bit rate=20Gb/s)

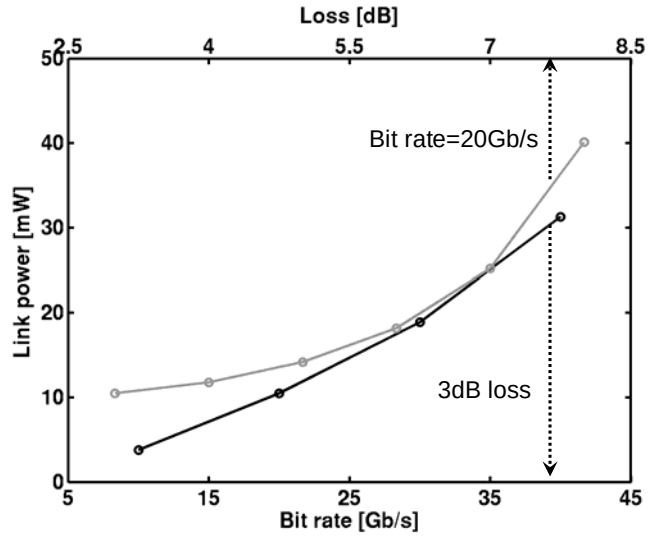


Fig.4.12 Impact of bit rate and link efficiency (optical power loss) on the optimum link power (C_{det}=C_{mod}=50fF and laser power (P_{laser})=1mW). The gray curve is the link power vs. loss at 20Gb/s bit rate. The black curve is the link power vs. bit rate at 3dB loss.

Similarly, reducing off-chip laser power reduces optical power at the receiver, which results in a larger receiver power. At the same time it decreases the static power at the transmitter. The impact of the laser power is similar to that of the loss as shown in Fig.4.13 since the increase in the receiver power is greater than the decrease in the static power at the transmitter.

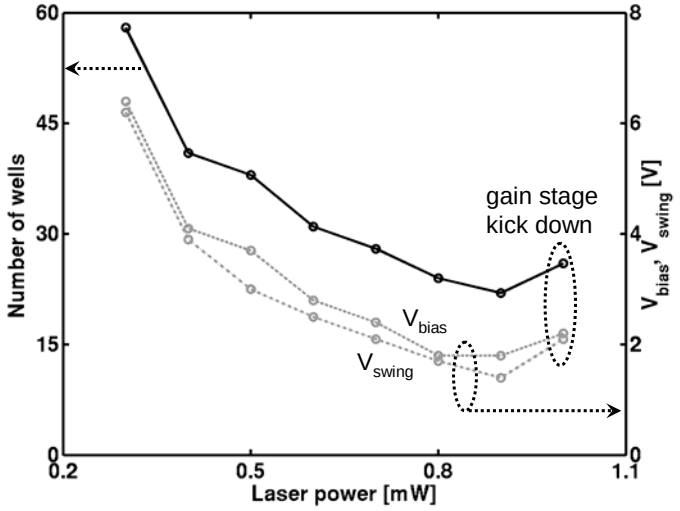


Fig.4.13 Impact of laser power (P_{laser}) on the optimum modulator design parameters: number of wells, V_{bias}, and V_{swing} (C_{det}=C_{mod}=50fF, loss=3dB, and bit rate=20Gb/s)

The detector and modulator capacitance are critical parameters in making optical interconnects viable for short distances [15]. Current flip-chip bonding process has about 50fF including the bump capacitance [10], but Ge-based CMOS compatible modulator [16] and MSM (Metal-Semiconductor-Metal) photo-detectors [17], [18] could lower capacitance to less than 10fF. Fig.4.14 shows the impact of detector/modulator capacitance on the optimization. With lower capacitances, both the dynamic power of the modulator and the receiver power decrease. The receiver power reduction is due to decrease in front-end noise with input capacitance. For 10fF and bit

rate=20Gb/s, the optimized swing voltage is less than the recommended supply voltage at 65nm technology node.

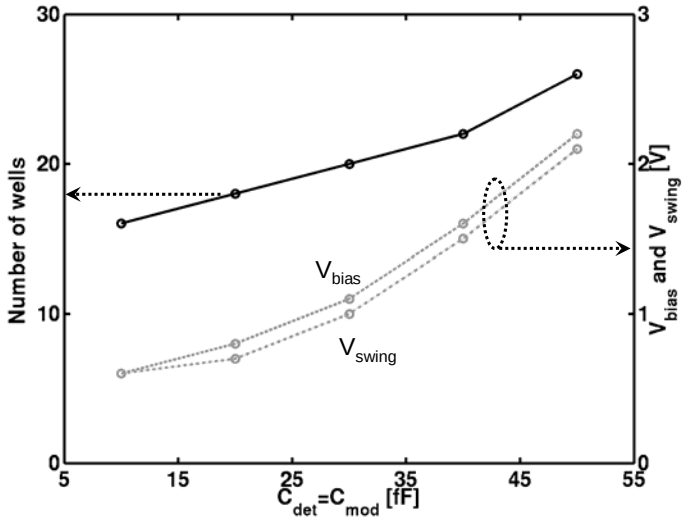


Fig.4.14 Impact of C_{mod} and C_{det} on the optimum modulator design parameters: number of wells, V_{bias}, and V_{swing} (laser power (P_{laser})=1.0mW, loss=3dB, and bit rate=20Gb/s)

Fig.4.15 presents the optimized link power as a function of the off-chip laser power and the modulator/detector capacitances. Reducing the capacitances yields low link power. This power is less than 5mW for C_{mod}=C_{det}=10fF at bit rate=20Gb/s and loss=3dB, which amounts to a 70% reduction compared to the case with C_{mod}=C_{det}=50fF.

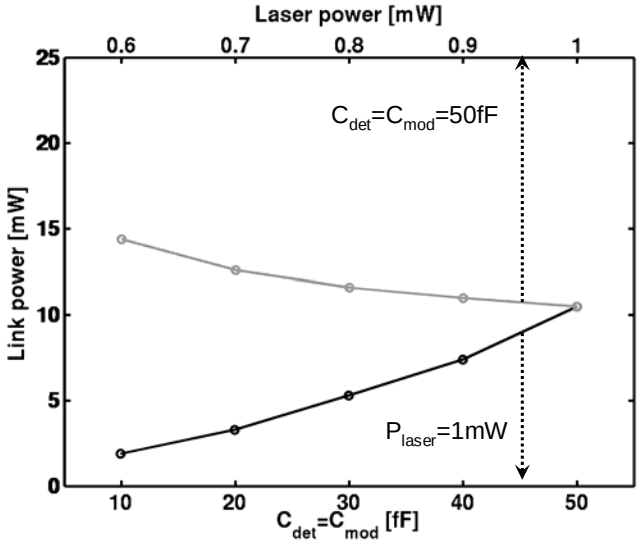


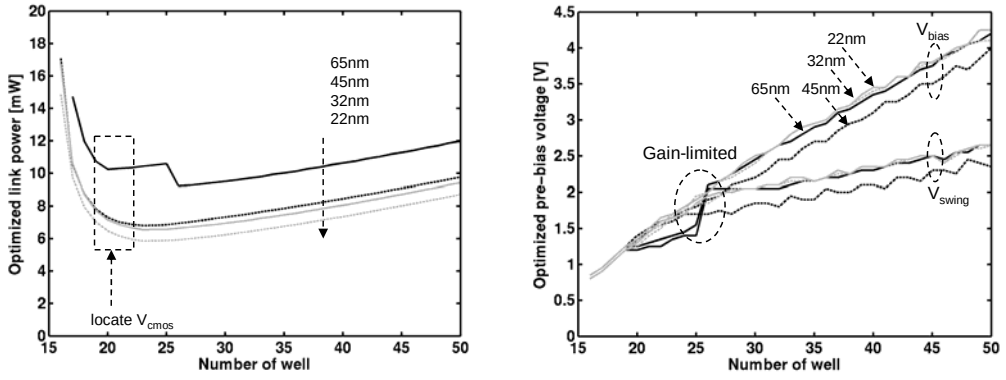
Fig.4.15 Impact of laser power and C_{mod} and C_{det} on optimized link power for bit rate=20Gb/s and loss=3dB. The gray curve is the link power vs. laser power at C_{det}=C_{mod}=50fF. The black curve is the link power vs. C_{det}=C_{mod} at 1mW laser power (P_{laser}).

4.5 Impact of Technology on Optimization

In section 2.5, we discussed the impact of technology scaling on the receiver and transmitter power for the specific design on the comparison between electrical and optical interconnects. It is observed that the receiver power dissipation is lower at future nodes as shown in Fig.2.15. The receiver power in the future is reduced because of improved transistors and lowering supply voltages; hence, it can change the optimum modulator design described in prior sections. In this section, we examine the impact of technology scaling on the optimization, focusing on the intermediate bit rate (up to 30Gb/s; 45nm technology node). For high bit rate (beyond 30Gb/s), we will

discuss the electroabsorption modulated lasers (EML) in the next section, which offers an attractive mean to provide high bit rate alleviating the package complexity.

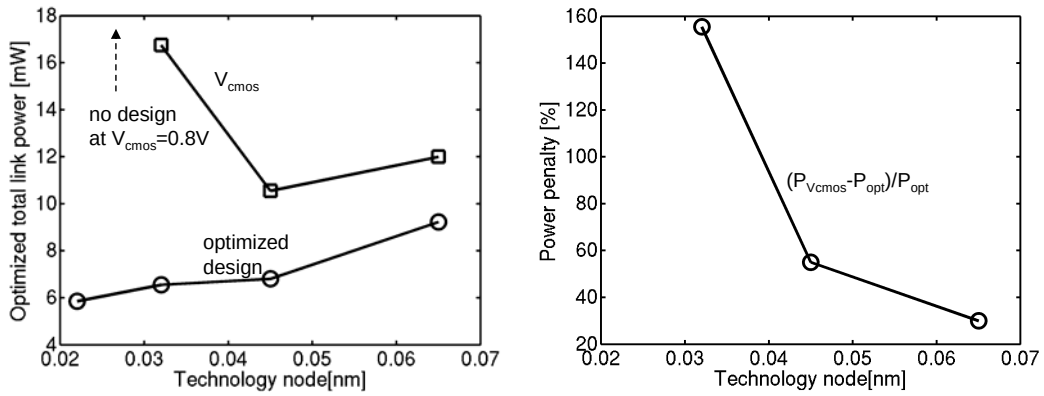
Fig.4.16 depicts the optimum modulator design metrics (number of wells, pre-bias and swing voltage) in terms of technology node scaling (transistor performance) for bit rate=20Gb/s and C_{mod}=C_{det}=50fF. The optimized link power decreases because of the receiver power, but the modulator design metrics are relatively insensitive to technology scaling because the required optical power at the receiver depends on the noise sources, which is determined by the photo-detector capacitance (C_{det}) and bit rate, but insensitive to transistor performance. It is observed that the optimized swing voltage (V_{swing}) is higher than the ITRS recommended supply voltage (V_{cmos}) for all technology nodes. This is because providing the required input optical power at the receiver for 20Gb/s needs higher V_{swing} of the modulator and high device capacitances dissipates large power at the receiver, thus, it renders the link power sensitive to V_{swing}.



(a) (b)

Fig.4.16 Modulator optimization in terms of technology scaling (transistor performance) (a) link power (b) pre-bias and swing voltage for bit rate=20Gb/s, C_{det}=C_{mod}=50fF, and P_{laser}=1mW

Consequently, operating modulator at V_{cmos} leads to increase link power as shown in Fig.4.17(a), depicting the link power operating at both optimized V_{swing} and V_{cmos}, for comparison. The gap increases as technology scales down, which, in turn, will result in higher power penalty as shown in Fig.4.17(b). This is because V_{cmos} decreases from 1.1V to 0.8V, as technology scales from 65nm to 22nm; the minimum V_{swing} becomes above V_{cmos} for 22nm technology node.

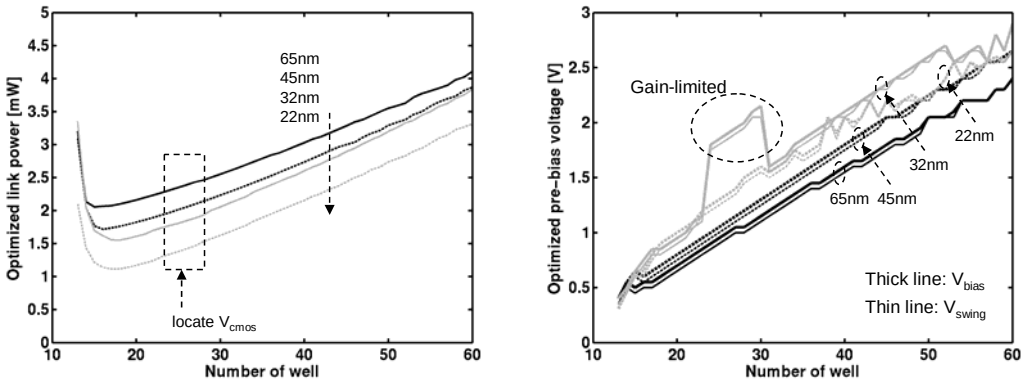


(a) (b)

Fig.4.17 (a) Link power for optimized swing voltage and ITRS CMOS supply voltage (b) power penalty of operating at the ITRS CMOS supply voltage for bit rate=20Gb/s, C_{det}=C_{mod}=50fF, and P_{laser}=1mW

Fig.4.18 and Fig.4.19 are the same plots with lower bit rate (10Gb/s) and small device capacitances (25fF). Lower bit rate, in essence, reduces the noise source of the front-end transistor, which, in turn, required input optical power at the receiver is reduced. Small detector capacitance also gives room for reducing the front-end transistor, possibly reducing the receiver power dissipation. Because of these two effects, the optimized V_{swing} of the modulator can be reduced, and possibly below V_{cmos} depending on the bit rate and device capacitances. The optimized modulator metric is

also insensitive to transistor performance as shown in Fig.4.18(b) because of the reasons aforementioned. There is the hump in the pre-bias and swing voltage around number of wells of 25 to 35. This is because the receiver goes into the gain-limited region, and higher pre-bias and swing voltage can reduce number of the amplifier gain stages, in turn, reducing receiver power. The power penalty, shown in Fig.4.19, has the opposite trend to the above one because the optimized V_{swing} becomes close to V_{cmos} as technology scales down.

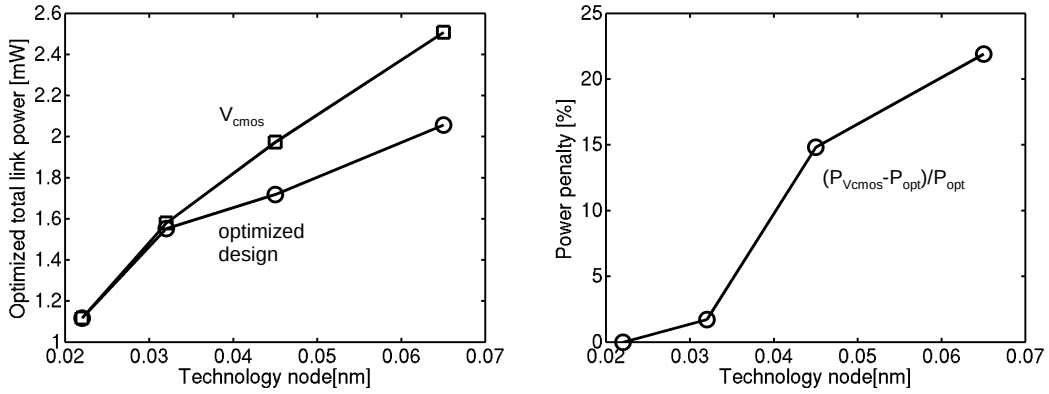


(a) (b)

Fig.4.18 Modulator optimization in terms of technology scaling (transistor performance) (a) link power (b) pre-bias and swing voltage for bit rate=10Gb/s, C_{det}=C_{mod}=25fF, and P_{laser}=1mW

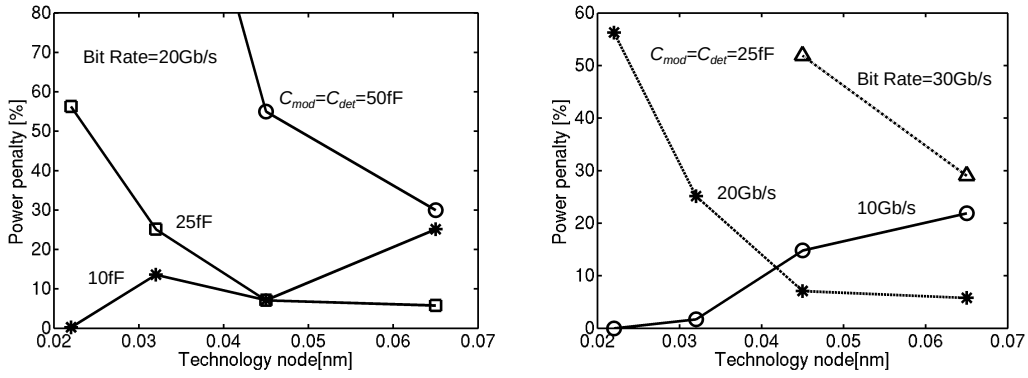
We observe that the power penalty has an opposite trend depending on the bit rate and device capacitances. Fig.4.20 depicts the power penalty in terms of bit rate and device capacitances including the two above cases. Between 10Gb/s to 20Gb/s and 10fF to 25fF, the trend of the power penalty changes to opposite one. For 45nm technology node, having the recommended bit rate 30Gb/s by ITRS, the power penalty becomes about 50% for device capacitances of 25fF. Beyond the 45nm technology node, higher bit rate demands high optimized V_{swing}. Since we expect that in future the

available optical power will increase because of stronger laser sources, it can give room for optimizing modulator metrics for beyond the 45nm technology node. Circuit solution, which pumps up the swing voltage using V_{cmos}, can also extend it, but this requires additional power dissipation, which, in turn, increases the power penalty.



(a) (b)

Fig.4.19 (a) Link power for optimized swing voltage and ITRS CMOS supply voltage (b) power penalty of operating at the ITRS CMOS supply voltage for bit rate=10Gb/s, C_{det}=C_{mod}=25fF, and P_{laser}=1mW



(a) (b)

Fig.4.20 Power penalty vs. technology nodes (a) in terms of device capacitances for bit rate=20Gb/s (b) in terms of bit rate for device capacitances of 25fF

4.6 Application: EML-based Optical Interconnect

As we discussed in section 2.5, an explosive growth in on-chip computational bandwidth demands a commensurate increase in off-chip I/O bandwidth as depicted in Fig.2.14. Beyond 32nm technology node, the demanded bandwidth becomes 30Gb/s and beyond, which is difficult to achieve with direct modulated VCSEL. Electroabsorption modulated lasers (EML), consisting of a continuous wave (CW) laser integrated with a passive transition element and an electroabsorption modulator (EAM) to provide modulation bandwidths in excess of 10Gb/s, were introduced a decade ago [19][20], and recently 40Gb/s EML was demonstrated [21]. The communication data rate possible using the EML exceeds the possible using commercially available directly modulated lasers for long-distance and intermediate-distance telecommunication applications. Higher level of integration (up to 12 independently addressable channels) have been achieved in data communication using parallel optical arrays of GaAs-based VCSEL operating at 10Gb/s [22][23]. However, such devices are currently limited to short-distance application, can not provide for sophisticated functions such as OEO conversion, and require large area. For dense wavelength division multiplexing (DWDM) application, EML offer an attractable mean even for short-distance application with higher bandwidth that VCSEL can not achieve or be achievable with parallel arrays with at least area penalty. In this section, we investigate the feasibility of EML, integrating VCSEL as a CW lasers and QWM, in terms of total optical link power and the impacts of bandwidth and device capacitances on the modulator optimization.

The electrical power dissipation of the CW laser used (3.3), replacing modulation current to DC current to provide the input laser optical power. The device parameters also are assumed to be the same values as shown in Table 3.1.

Fig.4.21(a) depicts the optimized link power as a function of input laser optical power with different bit rates for C_{mod}=C_{det}=50fF. For low bit rate (10Gb/s), the laser optical power has the optimum (about 0.7mW) minimizing link power, however, the laser power takes about 36% of the link power. The laser power is independent of the bit rate because laser functions only as a continuous wave source. For higher bit rate (20Gb/s and 30Gb/s), link power decreases drastically for small input laser optical power, because it provides small reflectivity difference, increasing the receiver power, and dissipates greater power at the transmitter to provide large reflectivity difference as shown in Fig.4.21(c)~(e). After a certain laser optical power (circled in Fig.4.21 (b)), the link power becomes relatively flat because increased laser power is compensated by reducing the modulator power. Achieving larger laser optical power addresses more complicated laser design; hence, the input laser power starting to saturate total link power is set to the optimum. The portion of the laser power decreases to about 23% and 20% for 20Gb/s and 30Gb/s at the optimum laser optical power, respectively. It indicates EML becomes attractive solution to achieve higher modulation bit rate.

Fig.4.22 depicts the same plot for different device capacitances (C_{mod}=C_{det}=10fF) and higher bit rates (up to 60Gb/s). These parameters are for 32nm technology node, achieving device capacitances possibly with monolithically integrated Ge-based CMOS compatible modulator [16] and MSM photo-detectors [17] [18] and bit rate that recommended by ITRS. The trend is similar with the case of 50fF, and the % of laser power range from 50% to 30% for 20Gb/s and 60Gb/s, respectively. The % of laser power for this case is higher even at high bit rate because the absolute link power excluding laser power is relatively small due to low device capacitances.

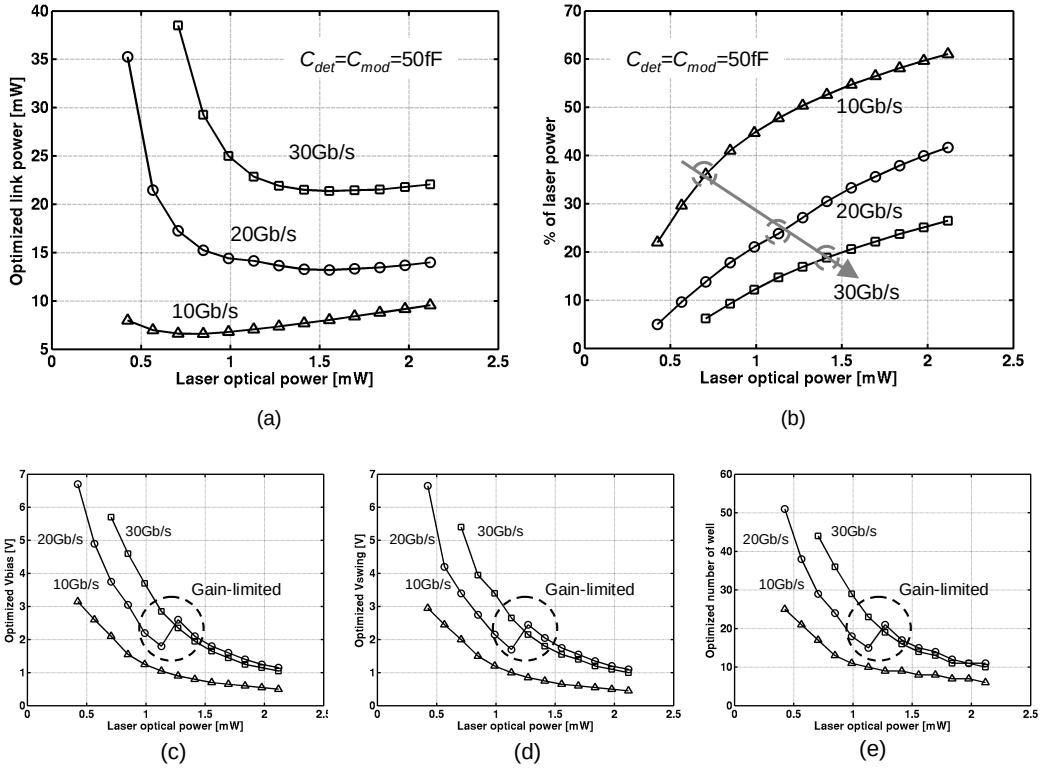


Fig.4.21 (a) Optimized link power (b) % of laser power in terms of input laser optical power (c)~(e) optimized modulator deign parameters with different bit rate at C_{mod}=C_{det}=50fF and assuming the loss from laser to modulator=1.5dB

Fig.4.23 presents the optimized link power and % of laser power as a function of input laser optical power with different device capacitances for 20Gb/s. As we discussed in two cases (50fF and 10fF), reducing the device capacitances decreases the laser power, which has a large fraction of the link power. For 10fF, laser power becomes more than 50%, while for 50fF, it decreases to about 23%.

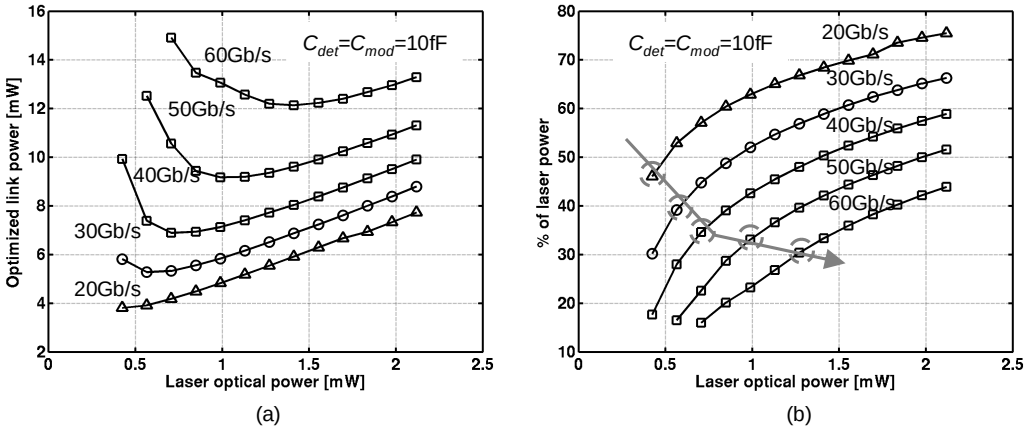


Fig.4.22 (a) Optimized link power (b) % of laser power in terms of input laser optical power with different bit rate at C_{mod}=C_{det}=10fF and assuming the loss from laser to modulator=1.5dB

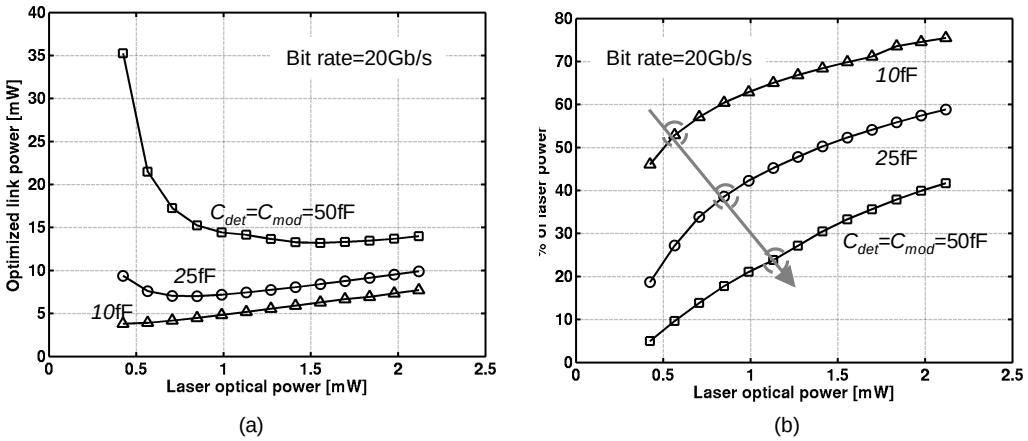


Fig.4.23 (a) Optimized link power (b) % of laser power in terms of input laser optical power with different device capacitances at 20Gb/s and assuming the loss from laser to modulator=1.5dB

4.7 Summary

We presented an optimization methodology for minimizing total optical link power, and obtained optimized modulator design parameters (number of wells, pre-bias voltage) and system parameters (swing voltage). As bit rate increases, the optimum swing voltage exceeds the stipulated supply voltage at 65nm technology node. This results in about 46% power penalty at 25Gb/s. A prohibitive power penalty for sub-optimum supply operation indicates a need for either another supply voltage for modulator driver or pumping circuitry to boost up the swing voltage. We also examine the impact of the device and system parameters (laser power, loss, bit rate, and capacitances) on both the optimum modulator design and operation parameters as well as the optimum link power. A higher bit rate results in an increase in the noise sources at the receiver. This, in turn, requires larger number of quantum wells in the modulator, which leads to a higher pre-bias and voltage swing. To mitigate this, either laser power can be increased or transmission loss can be reduced. However, both these factors have limits. An effective solution would be to reduce capacitances at the transmitter and the receiver. For example, reducing this capacitance to 10fF (achievable with Ge-based CMOS compatible modulator and MSM photo-detector) forces the optimum swing voltage to be within the stipulated supply voltage for 65nm technology node.

We also examine the impact of technology on the optimization and observed that the modulator design metrics are relatively insensitive to technology scaling (transistor performance) and power penalty has two opposite trend depending on bit rate and device capacitances. Lower bit rate and small device capacitances decrease the power penalty as the technology scales down; otherwise it has an opposite effect. For the 32nm technology node with considering recommended bit rate (30Gb/s), the power penalty becomes over 50%; beyond 32nm, a new circuit technique, pumping up

the voltage, and stronger laser power should be deployed to provide room for optimizing modulator metrics close to ITRS recommended chip environment.

For very high bit rate, over 30Gb/s recommended beyond the 32nm technology node, EML offers an attractive means of alleviating package complexity, sacrificing power dissipation. We examine the optimization laser optical power to minimize link power and investigate the fraction of laser power in terms of bit rate and device capacitances. For 32nm technology node, the laser power becomes about 30% of link power at the optimized condition.

4.8 References

[1] International Technology Roadmap for Semiconductors (2003), *http://public.itrs.net*.

[2] D. A. B. Miller, "Rationale and challenges for optical interconnects to electronic chips," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 728-749, June 2000.

[3] H. M. Ozaktas and D. A. B. Miller, "Limit to the bit-rate capacity of electrical interconnect form the aspect ratio of the system architecture," *Journal of Parallel Distributed Computing*, vol. 41, 1997.

[4] C. Svensson and G. D. Dermer, "Time domain modeling of lossy interconnects," *IEEE Transactions on Advanced Packaging*, vol. 23, no. 2, pp. 191-196, May 2001.

[5] W. J. Dally and J. Poulton, "Transmitter equalization for 4 Gbps signaling," *IEEE Micro*, pp. 48-56, 1997.

[6] M. Haycock and R. Mooney, *Proceedings of International Solid State Circuit Conference*, Feb. 2001.

[7] R. Mooney, "Scaling methods for electrical interconnects to meet the performance requirements of microprocessor platforms," in *14th Annual Workshop on Interconnects Within High Speed Digital Systems*, 2003.

[8] H. Y. Cho, P. Kapur, and K. C. Saraswat, "Power comparison between high-speed electrical and optical interconnects for inter-chip communication," *IEEE Journal of Lightwave Technology*, vol. 22, no. 9, pp. 2021-2033, Sept. 2004.

[9] D. T. Neilson, "Optimization and tolerance analysis QCSE modulator and detectors," *IEEE Journal of Quantum Electronics*, vol. 33, no. 37, pp. 1094-1193, July 1997.

[10] O. Kibar, D. A. A. Blerkon, C. Fan, and S. C. Easner, “Power optimization and technology comparison for digital free-space optoelectronics interconnects,” *IEEE Journal of Lightwave Technology*, vol. 17, no. 4, pp. 546-555, Apr. 1999.

[11] P. Kapur, R. D. Kekatpure, and K. C. Saraswat, “Minimizing power dissipation in optical interconnects at low voltage using optimal modulator design,” *IEEE Transactions on Electron Devices*, vol. 52, no. 8, pp. 1713-1721, Aug. 2005.

[12] K. W. Goossen, J. E. Cunningham, and W. Y. Jan, “Electroabsorption in ultranarrow-barrier GaAs/AlGaAs multiple quantum well modulators,” *Apply Physics Letter*, vol. 64, no. 17, pp. 1071-1073, 1994.

[13] A. V. Krishnamoorthy and D. A. B. Miller, “Scaling optoelectronic VLSI circuits into the 21st century: A technology roadmap,” Proc. *J. Sel. Topics Quantum Electronics*, vol. 2, no. 1, pp. 55-76, Apr. 1996.

[14] G. F. Williams, “Lightwave receivers,” *in Topics in Lightwave Transmission Systems*, T. Li. Ed. New York: Academic, 1991.

[15] H. Y. Cho, P. Kapur, and K. C. Saraswat, "Impact of Technology node on Power Comparison for High-speed Electrical and Optical Interconnects," *International Interconnect Technology Conference*, pp. 177-179, June 2005.

[16] Y. H. Kuo, Y. K. Lee, Y. Ge, S. Ren, J. E. Roth, T. I. Kamins, D. A. B. Miller, and J. S. Harris, “Strong quantum-confined Stark effect in germanium quantum-well structures on silicon,” *Nature*, vol. 437, no. 27, pp. 1334-1336, Oct. 2005.

[17] D. Buca et. al., “Metal-Germanium-Metal ultrafast infrared detectors,” *Journal of Applied Physics*, vol. 92, no. 12, pp. 7599-7605, Dec. 2002.

[18] C. O. Chui, A. K. Okyay, and K. C. Saraswat, “Effective dark current suppression with asymmetric MSM photodetectors in Group IV semiconductors,” *IEEE Photonic Technology Letters*, vol. 15, no. 11, pp. 1585-1587, Nov. 2003.

[19] Y. Kawamura, K. Wakita, Y. Yoshikuni, Y. Itaya, and H. Asai, *IEEE Journal Quantum Electronics*, vol. 23, no. 6, pp. 915-918, June 1987.

[20] H. Soda, M. Furutsu, K. Sato, N. Okazaki, Y. Yamazaki, H. Nishmoto, and H. Ishikawa, "High-power and high-speed semi-insulating BH structure monolithic electroabsorption modulator/DFB laser light source," *Electronic Letter*, vol. 26, pp. 9-10, 1990.

[21] [On-line] Infinera News available at http://www.infinera.com/news/2006-03-27.html.

[22] D. Kuchita, Y. Kwark, C. Schuster, C. Baks, C. Haymes, J. Schaub, P. Pepeljugoski, L. Shan, R. John, D. Kucharski, D. Rogers, and M. Ritter, "120Gb/s VCSEL-based parallel optical link and custom 120Gb/s testing station," in *Proceedings of 54th Electronic Components Technology Conference*, pp. 1003-1011, June 2004.

[23] J. Simon, L. Buckman Windover, S. Rosenau, K. Giboney, B. Law, G. Flower, L. Mirkarimi, A. Grot, C.-K. Lin, A. Tandon, G. Rankin, and R. Gruhlke, "Parallel optical interconnect at 10Gb/s per channel," in *Proceedings of 54th Electronic Components Technology Conference*, pp. 1016-1023, June 2004.

Chapter 4: QWM Design Methodology

Chapter 5

Design of Prototype Chip

5.1 Introduction

In the last three chapters, we quantified the power dissipation of both the electrical and optical links and examined the impact of device and system parameters on critical length defined as the interconnect length beyond which optics becomes more power efficient than their electrical counterparts. Having modeled these links, our objective in the second part of this dissertation is to verify these results in a real chip environment. The main design goal in a prototype chip which decouples the contribution of various link components on total power enables us to assess the impact of various figures of merit of devices (*CR*, *IL*, and coupling for optical and termination mismatch, crosstalk, attenuation, and package parasitics RLC for electrical links) at the system level. Toward this goal we implement optical and electrical design parameters which can be systematically controlled and varied using external electrical signals as discussed section 5.3.3.

We start with the chip configuration in section 5.2, showing the placement of each circuitry block, electrical and optical links' components, in a chip. This section gives the conceptual outline of the prototype optoelectronic chip and is followed by detailed circuit description. Section 5.3 deals with the electrical link, including signaling, clock generation, transceiver, on-chip termination resistor, and mode registers. In section 5.4, we discuss the optical link, similarly including signaling, transceiver with adjustable design parameters. Section 5.4 also deals with the fabrication process of the modulator and building optoelectronic chips using flip-chip bonding technique.

5.2 Chip Configuration

The main object of this prototype optoelectronic chip is implementing both electrical and optical links in a single chip for more a fairer comparison. For free space optical link, the modulator and photo-detector on a chip need to be exposed after packaging. The chip on board (COB) technique mounting the chip directly on the printed circuit board provides the solution, but the wire bonding is used for electrical signals rather than chip size package or ball grid array package, which increases the package parasitics due to long bonding wires. Fig.5.1 shows the fabricated prototype chip's photo, showing the placement of each circuitry blocks of electrical and optical links. The chip size is 5mm×5mm and has wire bond pads for electrical signals at the edge of the chip and flip-chip bond pads for optical devices (modulator and photo-detector) on the left center. This chip consists of electrical transceivers, optical transceivers, and common circuitry, such as phase locked loop (PLL), delay locked loop (DLL), mode register, and pseudo-random bit sequence (PRBS). The electrical link has unidirectional signaling transceivers (source synchronous clock) and

bidirectional signaling transceivers on the right side. Under the flip-chip bond pads, on the other hand, optical transceivers (transmitter on top, receiver on bottom) are placed with the clock distribution. The reference clock (low frequency) feeds to the phase locked loop, multiplies it and thus generate clean clocks (high frequency) for use in delay locked loop. DLL aligned the internal clock to the center of the received data to maximize the timing margin. Mode registers, on the bottom, controls the design parameters to examine the impact of these parameters on the comparison discussed in Chapter 2 and optimize the design parameters minimizing the optical link power. The detailed circuit description is discussed in the following sections.

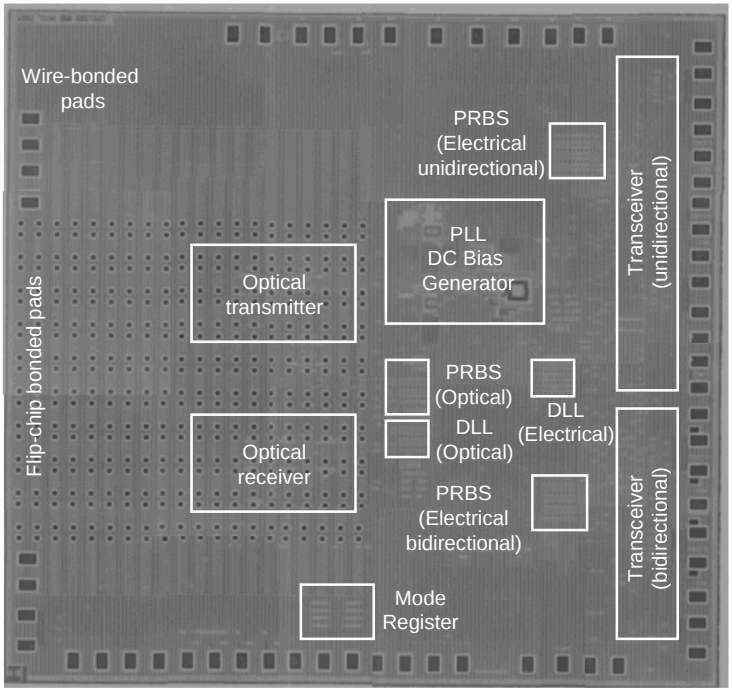


Fig.5.1 Chip configuration of prototype electrical and optical link

5.3 Signaling Scheme of Electrical Interconnect

Scaling in device structure and fabrication technologies decreases the cost of transistors faster than the cost of I/O pins, making pin saving an important cost parameter. Thus, signaling setups that reduces the number of pins and wires are attractive alternatives to the traditional unidirectional differential system. Differential unidirectional signaling can transmit data with the shortest bit period because this signaling minimizes the generation of noise and sensitivity to it. However, to maximize the total bandwidth available to chip, the bandwidth per pin should be maximized. Additional transistors combined with enhanced signaling can maximize the bandwidth per pin. Bidirectional signaling can double the bandwidth per pin by using each pin to both transmit and receive data. However, this signaling increases both the generation of noise and sensitivity to noise. To alleviate the noise effects, we keep the differential signaling with the current mode channel [1].

Fig.5.2 shows the interface architecture of source synchronous point-to-point bi-directional point-to-point parallel link, and the timing of the corresponding interface signals. There are three primary components in a link: the transmitter, the channel, and the receiver. The transmitter converts a digital signal data sequence into an analog signal in the channel; the channel is the entire transmission path along which the signal travels; and finally the receiver converts the received analog signal back to digital data sequence. Transmission of all data signals, $data_0[0\text{-}n]$, and a reference clock signal, $refClk_0$, is triggered synchronously (hence the name source synchronous) by $TxClk_0$. The receiver timing recovery circuitry generates a global receiver clock, $RxClk_1$, by delaying the received clock, $refClk_0$, by half of a bit time. $RxClk_1$ is then used to sample all incoming data signals in the middle of their transitions to maximize timing margins.

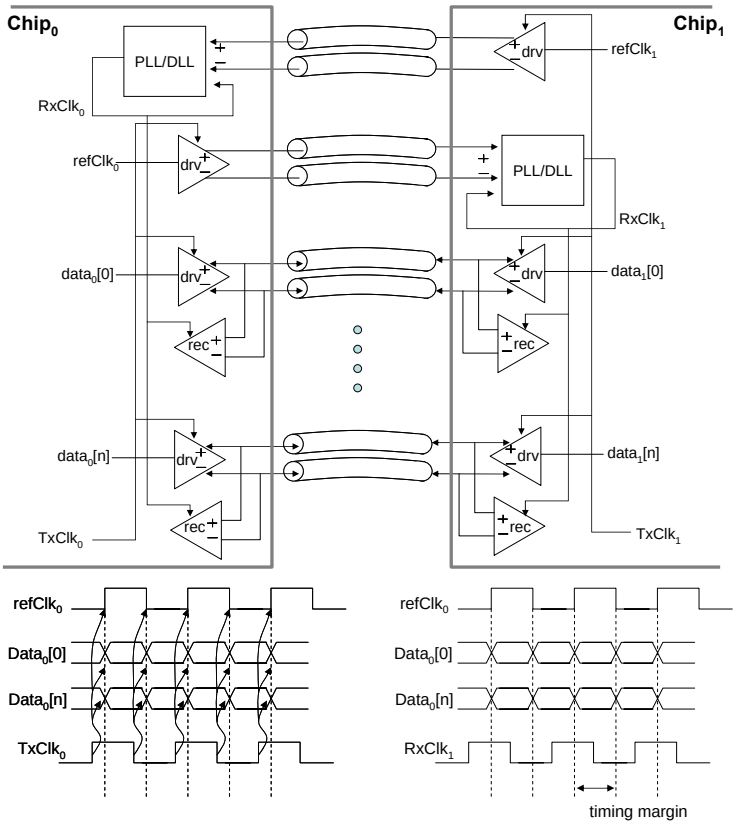


Fig.5.2 Schematic and I/O timing diagram for source synchronous bi-directional and differential point-to-point parallel link

5.3.1 Clock Generation

The skew and jitter in the local receiver and transmitter clock (*RxClk*, *TxClk*) followed by clock distribution greatly affect the timing margin in the received signals. These factors depend largely on the performance of the DLL used to generate *RxClk* and *TxClk*. We used the dual-loop architecture DLL, which allow a much wider locking range than the single loop DLL [2]. Fig.5.3 shows the clock generation and

distribution, using external low frequency clock (125MHz). The received *refClk* can be noisy and jittery. If it is fed directly to the delay line input of the delay lock loop (DLL), the jitter is passed to the output clock of the DLL, making the output worse than the input. To alleviate the problem, the delay line takes the separate clean clock source as input. Phase locked loop (PLL) generates noise suppressed clocks, multiples of external clock (0.625GHz~5GHz). Clock select (*selClk*) generated from mode registers, which is discussed in section 5.3.3, selects one of these clocks, and feeds to the input of the DLL.

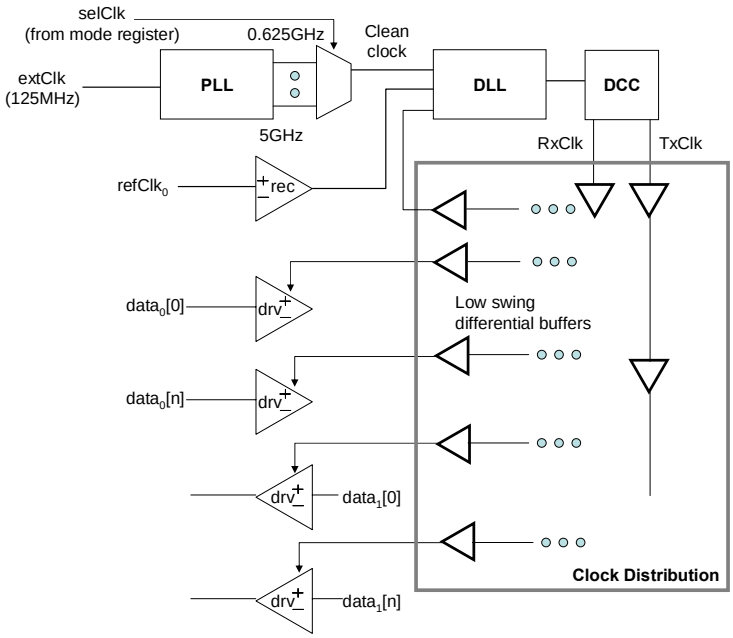


Fig.5.3 Clock generation and distribution

Although low swing differential clock buffers dissipate more power than using CMOS inverters and the area of each low swing buffer is also much larger than a CMOS inverter of comparable strength, jitter of the local *RxClk* and *TxClk* can be substantially reduced, providing receivers more timing margins. This is more crucial for our prototype chip because electrical and optical transceivers are in a single chip

and widely spread out, which results in widely distributed clocks.

5.3.2 Transmitter and Receiver

5.3.2.1 Data Generation and Data Path

To test the functionality, the links should operate correctly for any random data pattern. We implement the pseudo random bit sequence (PRBS) using 31 bit maximum-length linear feedback shift register (LFSR) chain: an M-sequence with primitive function

$$f(x) = x^{31} + x^3 + 1 \tag{5.1}$$

and repeating period 2^{31}-1 bits. Standard CMOS logic (flip-flop) considered local interconnect RC loading becomes marginal to implement the dual edge clock. Thus, Dual edged flip-flop with high-speed CMOS current mode logic (CML) [3] is used for reliable operation at the high frequency.

Fig.5.4 shows the block diagram of the data path, whereby PRBS generates the random data to the transmitter and detects the bit error rate in the received signal. Optical and electrical link have its own data path (PRBS generator and verifier) because the clocks are aligned using different DLL. After PRBS verifier, muxes controlled by *sel[1:0]* choose one of error signals from the PRBS verifier, enabling the separate test of bit error rate with a single pin (error).

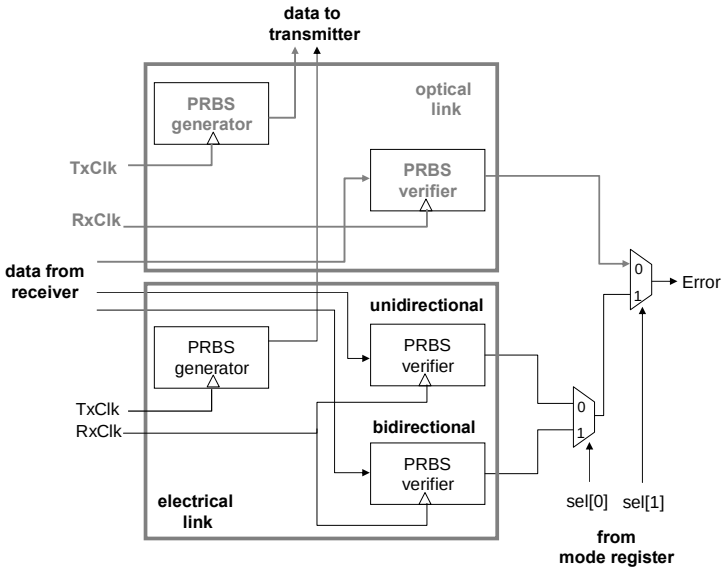


Fig.5.4 Block diagram of the data path

5.3.2.2 Transmitter: Bipolar Differential Current Mode Driver with Equalization (Pre-emphasis)

A transmitter circuit encodes a symbol into a current (current mode) for transmission over a line. Current mode transmitters have a very high output impedance and are implemented using transistors operated in saturation. High impedance of current mode signaling enables noise immunity to power supply. Using source coupled pair to steer the current from the current source into one of two legs, as shown in Fig.5.5. The symmetric nature of bipolar signaling gives roughly equal transmitter offset for the “0” and “1” symbol. Unipolar signaling, on the other hand, lumps most of its total offset into the “1” symbol, and thus, must reduce its detection threshold to balance its high and low noise margins. Bipolar signaling also reduces power dissipation because for a given signal swing, a bipolar system draws half of peak current form the supply as a unipolar system. However, bipolar signaling pulls up and

down the signal lines, hence it requires additional area penalty. The coupling capacitances (C_{in}^{+} and C_{in}^{-}) degrade the output signal slope. The compensate capacitors connected to reverse inputs compensate the degradation.

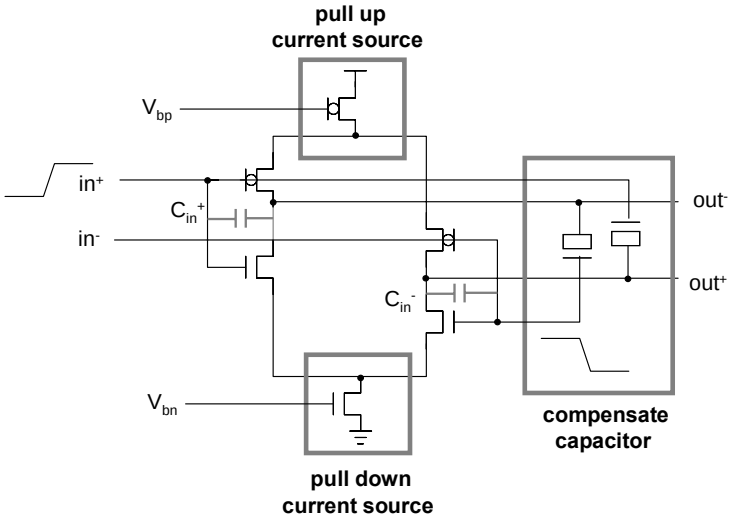


Fig.5.5 Differential bipolar current steering driver

The resistance of long off-chip lines increases as the square root of signal frequency because of the skin effect, resulting in inter-symbol interference as unattenuated low frequency signal components overwhelm the high frequency components. This effect can be countered by equalizing the transmission line, amplifying the high frequency components, or attenuating the low frequency components to level the frequency response. Pre-emphasis transmitter has a two-tap symbol-spaced finite-impulse response (FIR) filter that is used to cancel the trail of the transmission line response of two subsequent symbol intervals [4] as shown in Fig.5.6.

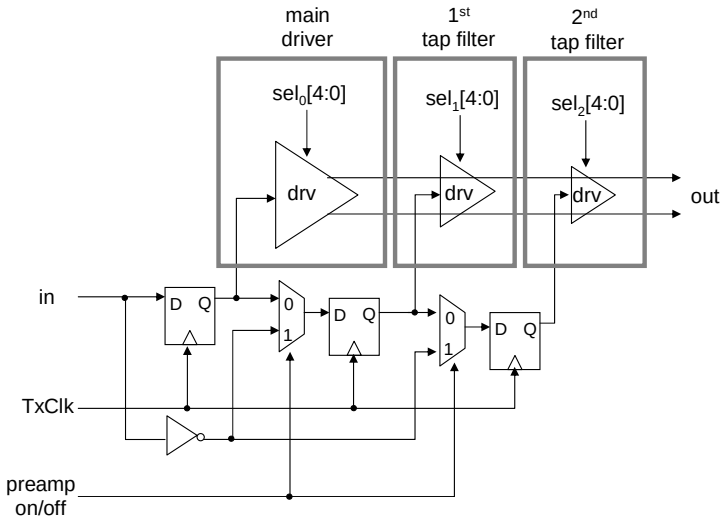


Fig.5.6 Pre-emphasis transmitter with two-tap FIR

5.3.2.3 Receiver: Four-input Differential Amplifier

A receiver detects an electrical quantity, current or voltage, to recover a signal from a transmission medium. A receiver should have good resolution and low offset in both the voltage and timing dimensions. Fig.5.7 shows the receiver for differential simultaneous bidirectional signaling, four-input differential amplifier [5]-[7]. *Vfr* is summed of forward traveling wave (*Vf*) and reverse traveling wave (*Vr*). The replica driver (1/4 strength of main driver) generates a replica of the forward traveling wave (r^+ and r^-). The four-input amplifier subtract this differential replica from the differential line voltage to recover the reverse traveling wave (*Vrr*). Differential simultaneous bidirectional signaling requires fewer pins than unidirectional single-end signaling, because fewer signal returns are needed, and largely eliminates signal-return cross talk because the currents in both directions are balanced.

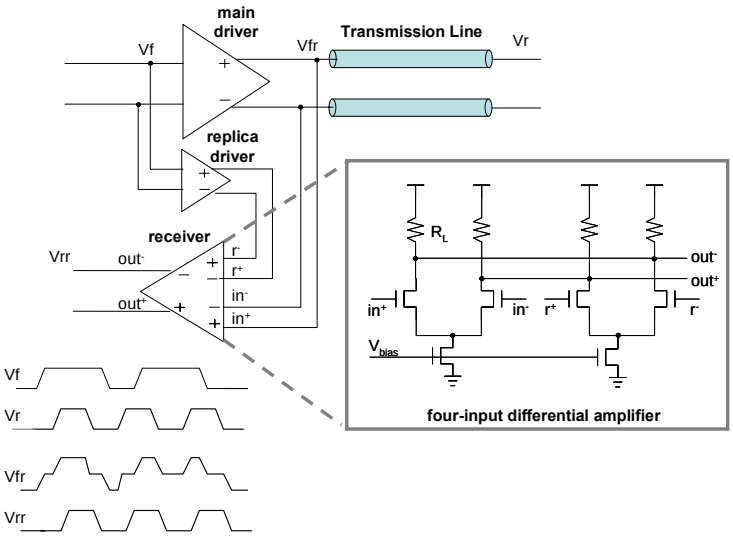


Fig.5.7 Receiver for differential simultaneous bidirectional signaling

5.3.2.4 On Chip Mid-supply Termination

The impedance discontinuities bounce back traveling waves and reflected waves and thus is the cause of inter-symbol interference. Resistive terminations are always required in high performance signaling systems to absorb traveling waves, preventing unwanted reflections, for incident wave switching. Signal lines can be terminated either at the transmitter or at the receiver and in some signaling systems, terminations are provided at both the transmitter and the receiver. In high speed signaling, it is often advantageous to use on-chip termination resistors rather than off-chip discrete resistors. Off-chip termination always results in an un-terminated stub composed of the package parasitics and the internal circuitry, and this stub generally introduces unaccepted large reflections into the signal line [1]. Bonding the chip directly on the board (COB) with long bonding wires makes the chip more vulnerable because of large package parasitics. The bipolar mode signaling needs to set the common mode voltage on the signal line to the mid-supply voltage. More importantly,

to assess the impact of impedance mismatch between the transmission line on PCB and on-chip termination resistor, the impedance of terminator need to be adjustable.

Fig.5.8 shows the circuit diagram of the termination resistor based on common-gate configuration [8]. The input impedance looking into the sources of two common-gate configurations is used for termination resistor. The external resistor (R_{EXT}) is used as a reference to set the bias current and termination resistance. The resistor-based voltage divider (R_1~R_4) generates three reference voltages, V_H, V_M, and V_L. The two biases (V_H and V_L) set the termination voltage and impedance by providing bias voltages, V_P, and V_N. The current mirrors flow the same current of $I_B=V_M/R_{EXT}$. M7 and M9 are replicas of M3, and M8 and M10 are replicas of M4.

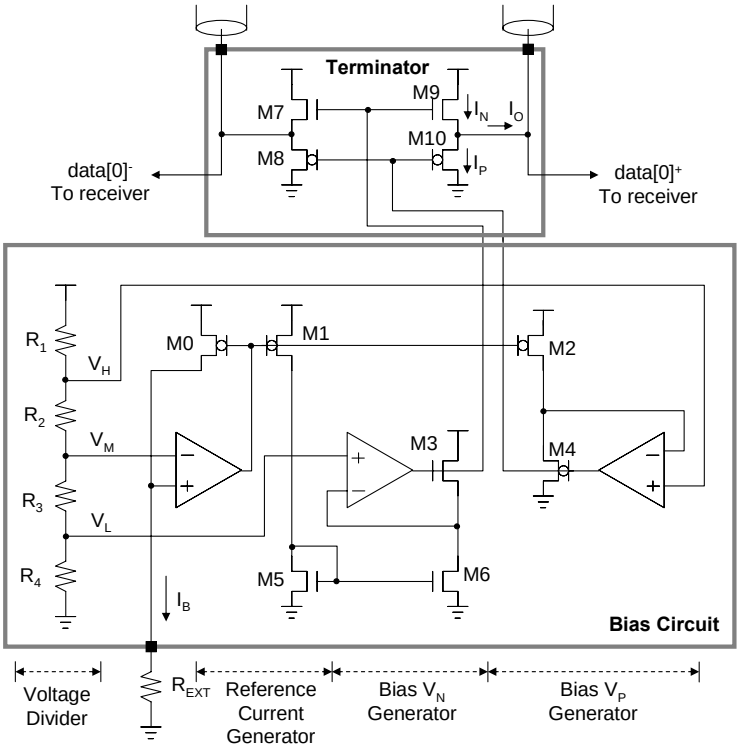


Fig.5.8 Circuit diagram of the termination resistor and bias circuit

Fig.5.9 shows I-V characteristic for the mid-supply on-chip termination resistor. Although the source impedance of neither PMOS nor NMOS is linear, the combined effect of PMOS and NMOS results in close to linear resistance around the termination voltage. Assuming ideal square-law current characteristic of the MOS transistors, the output current is given by

$$I_O = I_P - I_N = K_P(V_O - V_P - |V_{TP}|)^2 - K_N(V_N - V_O - |V_{TN}|)^2 \tag{5.2}$$

where K_P and K_N are the transconductance parameters and V_{TP} and V_{TN} are the threshold voltages of PMOS and NMOS respectively. When we set the reference voltage (V_H and V_L) such that only one of the output transistor is on and the other is just on the threshold, with $K_P = K_N$, the output current I_O is given by

$$I_O = \frac{1}{R_T}(V_O - V_M) \tag{5.3}$$

where

$$R_T = \frac{V_H - V_L}{2 \cdot I_B} \text{ and } V_M = \frac{V_H + V_L}{2} \tag{5.4}$$

From the bias condition of $I_B = V_M / R_{EXT}$, the termination resistance is given by

$$R_T = R_{EXT} \frac{V_H - V_L}{V_H + V_L} \tag{5.5}$$

and thus, the termination resistance is adjustable by R_{EXT}. The derivation is based on

the assumption of ideal square-law current characteristic of long-channel MOS transistors. However, nonideal effects such as short-channel effects, process and voltage variation, deteriorate the linearity of termination resistance. Under a typical condition, the maximum deviation is 5.5% over a full voltage range [8].

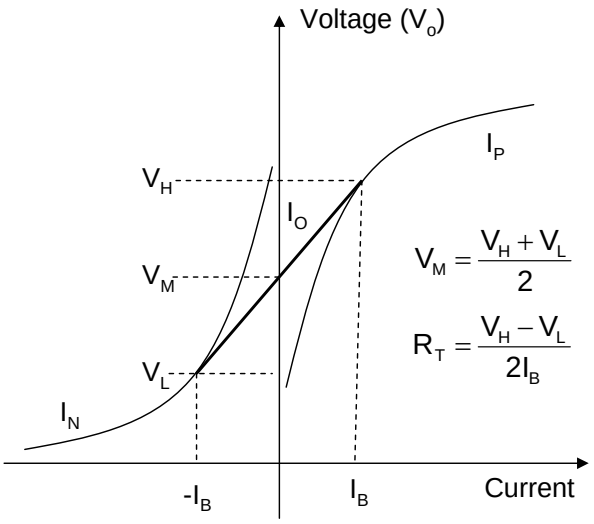


Fig.5.9 I-V curve of the on-chip termination resistor

5.3.3 Mode Register

Adjusting design parameters is one of the key design goals, to examine the impact of these parameters on the power comparison discussed in Chapter 2 and 3. For optical link, this scheme is also used to optimize the design parameters minimizing total link power dissipation. Fig.5.10 presents the block diagram and timing of mode register set, comprising of *en* (enable), *clk*, and data. Before main testing cycle, the serial registers are set during the mode register set cycle as shown in Fig.5.10, which performed at the very low frequency (~KHz). The output of serial registers chooses the designated design parameter set, controlling passing gate, and adjusts the design

parameters.

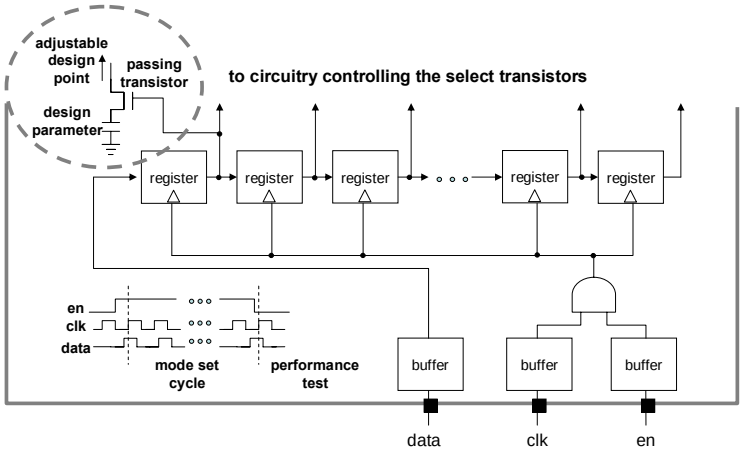


Fig.5.10 Block diagram and timing of mode register set

5.4 Optical Interconnect: Process and Optoelectronic chip

Optical interconnect implemented in this chip uses source synchronous signaling similar with electrical counterpart discussed in section 5.3. However, optical interconnect is fundamentally unidirectional and unipolar signaling in a single-ended signal line as shown in Fig.2.1. Fig.5.11 shows optical link using free space as a transmission medium and modulator-based transmitter. Off-chip laser is focus on the flip-chip bonded modulator (discussed in section 5.4.2); modulator modulates the light by modulator driver on CMOS chip; light propagates to the flip-bonded photo-detector by mirrors; photo-detector converts the light to the current; following receiver converts this current to the voltage, then gain stages amplify to the swing voltage. In this section, we discuss the design of the transmitter and the receiver and the process building optoelectronic chips using flip-chip bonding.

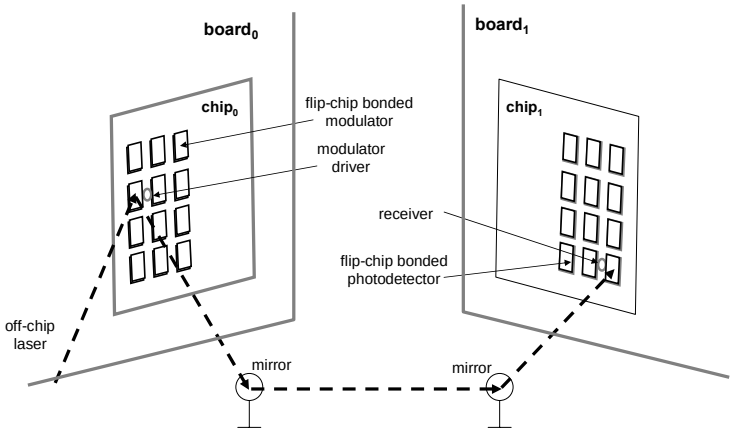


Fig.5.11 Schematic of the optical link

5.4.1 Transmitter and Receiver

5.4.1.1 Transmitter

The figure of merit of the modulator, contrast ratio (*CR*), depends on the swing voltage of the modulator driver as discussed in section 3.5 and optimized swing voltage depends on the bit rate, greater than supply voltage (V_{dd}=1.2V) of chip technology node (90nm). This voltage is simply driven by the CMOS inverter chains as shown in Fig.5.12 with the different voltage (V_{ddm}). To convert from supply voltage (V_{dd}) to modulator swing voltage (V_{ddm}), we use the level shifter, and it degrades duty ratio. In this design, for V_{ddm}=1.5V and 10Gb/s (100ps signal), 2% (2ps) of duty distortion is obtained with HSPICE simulation.

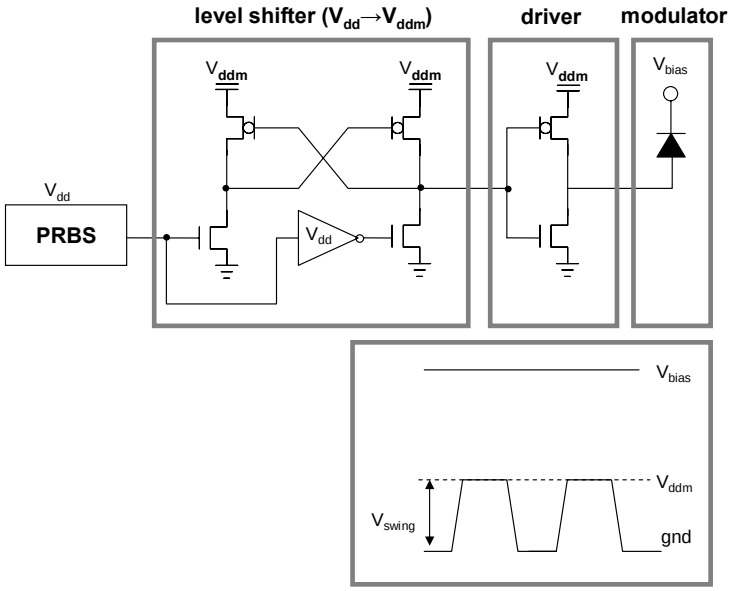


Fig.5.12 Transmitter: quantum-well modulator-based optical link

5.4.1.2 Receiver : Transimpedance Receiver

The transimpedance amplifier is the most widely used preamplifier structure for high-speed optical receiver [9]. Its merit is to combine a relatively high transimpedance gain with high speed. In an efficient high-speed design, the dominant pole is located at the input because of photo-detector capacitance and its processing capacitance (e.g. bump capacitance). Fig.5.13 shows the transimpedance receiver with adjustable design parameters, such as input capacitance, front-end receiver size, and feedback resistance. Input capacitance and front-end receiver size are adjusted with passing NMOS transistors controlled by *csel[3:0]* and *msel[3:0]* respectively, which are set by mode register cycle. However, feedback resistance is adjusted by gate bias of the PMOS transistor, which is driven by an external pad (V_{tia}). By adjusting these design parameters, we could find the optimized the design minimizing total optical link power as discussed in section 2.2.

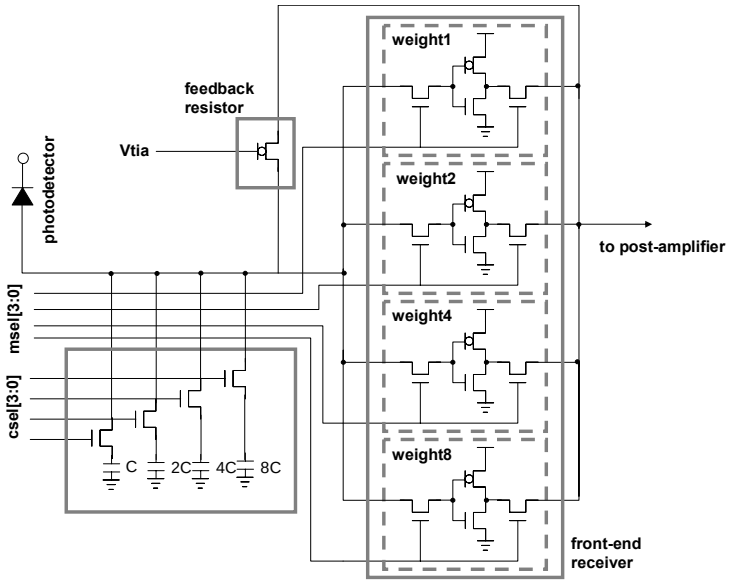


Fig.5.13 Transimpedance receiver with adjustable design parameters

5.4.1.3 Receiver: Offset-tolerant Replica Biasing with Accurate Threshold Control

To obtain a completely integrate optical receiver, the preamplifier's output signal is further amplified to rail-to-rail voltage. A post-amplifier, based on a string of modified inverters biased at their threshold voltage, is used, whereby the first inverter have a small signal and act as linear amplifiers and at the end of the chain, clipping occurs at the ground and power-supply voltage. The advantage of this approach is that the exact serial number of the inverter where the clipping begins is irrelevant. A large signal simply results in a shift forward of the first clipping. This results in a large dynamic range without automatic gain control. The major challenge for the practical realization of the chain is the correct biasing of the inverters. It is performed with an offset-tolerant replica biasing circuit with accurate threshold control [10]. The replica

feedback loop consists of a level shifter, an inverter chain (M3-M5) in the signal path, low pass filter, a replica of the inverter M1 (M3), and a comparator alongside as shown in Fig.5.14. The dc bias voltage at the input of the first out-of-the-loop inverter (M1) is compared with the replica's threshold and forced to this voltage by adjusting the level shifter. Due to various effects of process and temperature variation, there is an offset, but the signal has been amplified by inverter chain (M3-M5), hence the standard deviation on this biasing error reduces by its total gain.

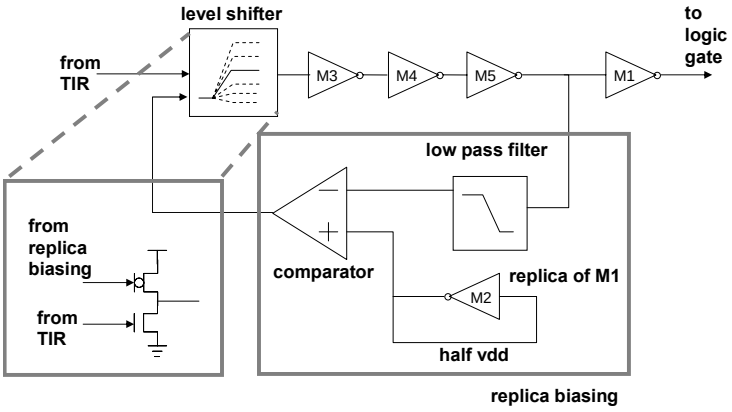


Fig.5.14 Post-amplifier with offset-tolerant replica biasing circuit

5.4.2 Modulator and Photo-detector

To make an electroabsorption modulator using a quantum-confined Stark effect, the typical approach is to place an absorbing quantum well in the intrinsic layer of a p-i-n diode. By applying a static reverse bias across the diode, photo-generated carriers are sufficiently swept out of the intrinsic region and the device acts as a photo-detector. Varying this bias causes a modulation in the optical absorption, resulting in an optical modulator. Surface-normal operation of MQW modulator is an effective way to achieve to high device density and simple optical coupling required for dense

optical interconnections. To increase the absorption of the photon in the device, surface-normal modulator often uses an absorbing region that contains many quantum wells. Quantum well modulators can be integrated to electronics in two-dimensional arrays by flip-chip bonding as discussed in section 5.4.3. Because the device can operate both as modulators and photo-detectors, fabrication and integration of a single device array adds both optical input and output capabilities to an electronic chip.

The growth of epitaxial structures by solid-source molecular beam epitaxy (MBE) is a well-known technique [11]. In order to use exiting diode lasers (with an output wavelength of 850nm) the GaAs well width was chosen to be 95Å and the $Al_{0.3}Ga_{0.7}As$ barrier thickness was 30Å. A design with 50 wells would achieve a satisfactory trade-off between exciton shift and total absorption [12]. Table 5.1 shows the final wafer design. The GaAs buffer layer is used to protect n-AlGaAs layer after integration and etch tuning step removal.

Table 5.1 Design of the modulator epitaxial structure

Description	Material	Thickness (Å)	Dopant (cm^{-3})
p cap layer	GaAs	100	$Be=1.0\times10^{19}$
p layer	p-$Al_{0.3}Ga_{0.7}As$	2030	$Be=1.0\times10^{19}$
i MQW layer	50× GaAs $Al_{0.3}Ga_{0.7}As$	95 30	
n layer	n- $Al_{0.3}Ga_{0.7}As$	5000	$Si=4.4\times10^{18}$
buffer layer	GaAs	500	
etch stop layer	$Al_{0.3}Ga_{0.7}As$	2800	
undoped (100) GaAs substrate			

5.4.3 Process: Building Optoelectronic Chip

Monolithic hybrid integration could give the highest performance by reducing parasitic capacitances [13]. However, flip-chip bonding, an approach common in electrical packaging, has several advantages. Flip-chip bonding substantially reduces the length of the electrical connections compared to the current commercial technique using short bond wires and carefully designed circuit board traces, thereby lowering both capacitances and inductances, and thus allows high-speeds and a reduction in power dissipation. Additionally, it enables entire device arrays to be integrated in a single operation, perhaps on a wafer-scale [14].

The detailed process step is as follows: 1) Circuit and device arrays are fabricated separately. 2) The contact pads of one or both chips are metalized to create diffusion barriers and the appropriate materials to act as a solder. Typical flip-chip bonding processes employ indium solders that are bonded to gold [15]. 3) The device array and CMOS chip are aligned in a flip-chip bonder and bonding is performed, generally by pressure and elevating the chip temperature to soften and melt the solder. 4) Following bonding, it is common to remove the III-V substrate by using selective wet or dry etch process, which remove the open opaque substrate as shown in Fig.5.15. Fig.5.16 shows the photograph of the fabricated optoelectronic device array after flip-chip bonding and substrate removal process (step (6) in Fig.5.15).

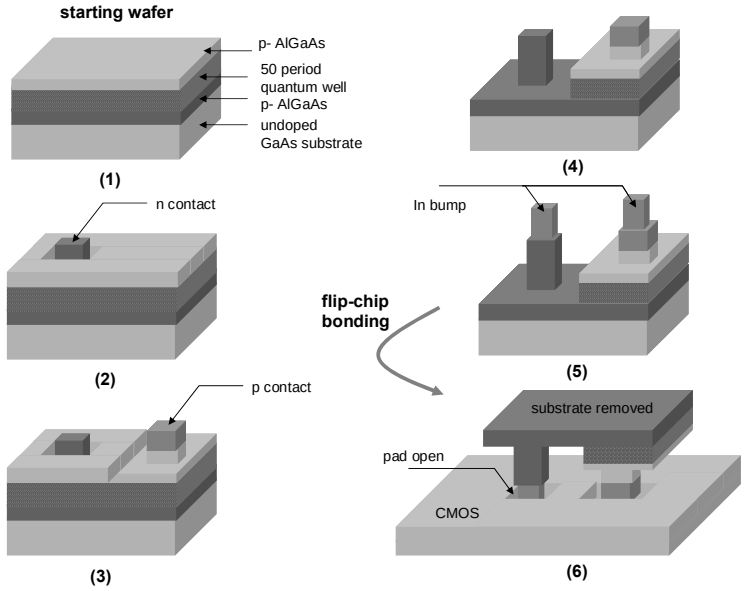


Fig.5.15 Fabrication process steps of optoelectronic chip: (1) starting quantum well wafer (2) wet etching n-well followed by n-contact (3) dry etching p-well after p-contact (self masking) (4) dry etching n-p-well defining standalone quantum well modulator (5) evaporating In bump for flip-chip bonding (6) substrate removal followed by flip-chip bonding, thus achieves an optoelectronic chip

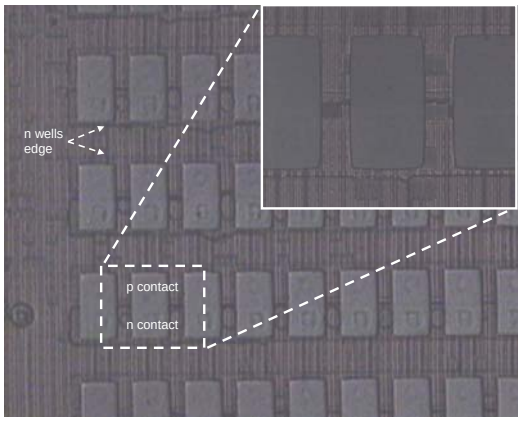
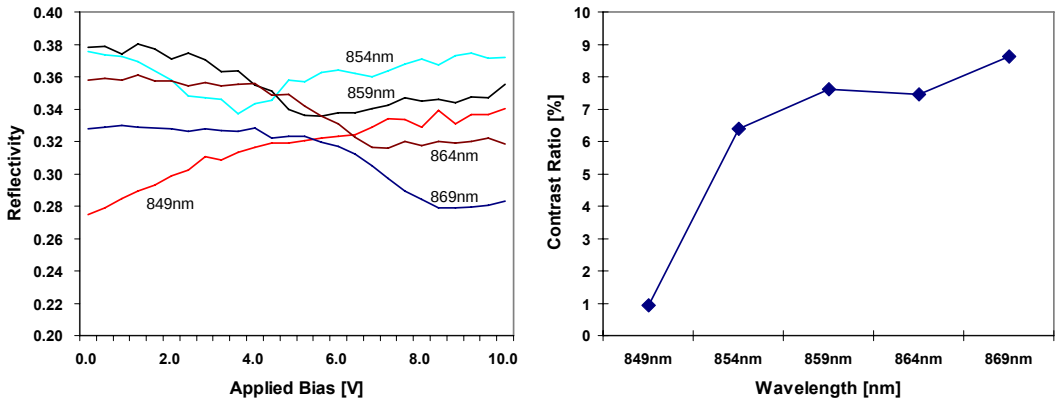


Fig.5.16 Photograph of the optoelectronic device array after flip-chip bonding and substrate removal process

5.4.4 Performance of Integrated Devices

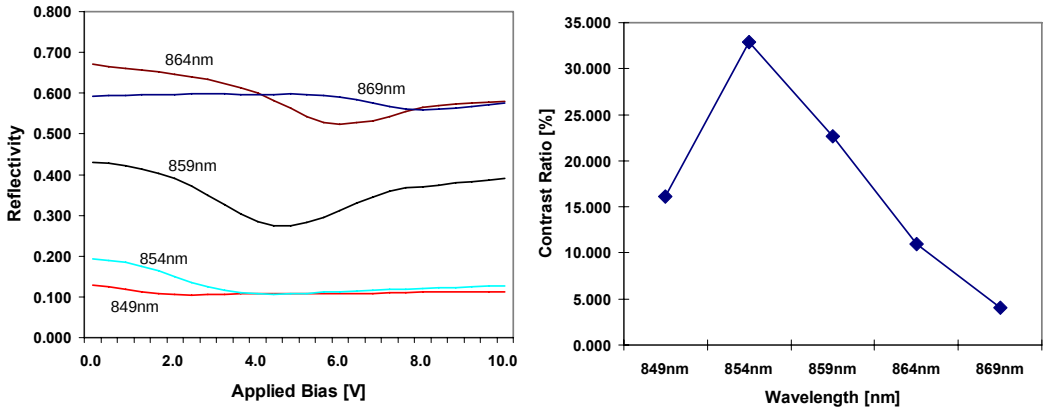
As we discussed in Chapter 4, the insertion loss and contrast ratio at a given voltage swing are the important metrics determining the performance, especially power dissipation of the optical links. Fig.5.17(a) shows the measured device reflectivity as a function of the swing voltage for various wavelengths. The very low reflectivity is due to additional losses that result from imperfect contact reflectivity (because of metallic diffusion occurring at the Au/GaAs interface), scattering at the top surface of the device, and reflective losses that occur because the devices were not anti-reflective (AR) coated. This lack of AR-coating created an optical cavity whose resonance was situated at a random wavelength [12]. Absorption is enhanced at the cavity resonance position, and weakened for off-resonance wavelengths, significantly changing the performance of the device. The measured contrast ratio of the device is only about 7% because of the afore-mentioned reasons.



(a) (b)

Fig.5.17 (a) Device reflectivity before cavity tuning (b) maximum contrast ratio at the voltage swing of 1.2V for various wavelengths

To obtain a working device (reasonable contrast ratio), the cavity resonance must overlap the exciton absorption peak. The cavity tuning procedure was performed following etch stop removal. A series of tuning cycles was employed, allowing more precise control of cavity resonance. A single tuning cycle consists of 15-second dips into hydrogen peroxide and then hydrochloric acid and water (1:1 ratio). The process works by oxidizing a thin layer of GaAs and subsequently stripping it off with hydrochloric acid. Fig.5.18(a) shows the measured device reflectivity as a function of the voltage swing for various wavelengths after 25 tuning cycles. The moving of the resonance peak is about 0.75nm per tuning cycle, which is consistent with an etch depth/cycle of ~20Å [12]. The measured contrast ratio improved significantly from 6.5% to 34% at 854nm and voltage swing of 1.2V.



(a) (b)

Fig.5.18 (a) Device reflectivity before cavity tuning (b) maximum contrast ratio at the voltage swing of 1.2V for various wavelengths after 25 tuning cycles

The specific voltage swing, 1.2V, is the typical CMOS voltage of the 90nm technology node. However, as we discussed in Chapter 4, the optimal voltage swing minimizing the link power dissipation is determined by device and system parameters,

and has a significantly higher voltage than CMOS voltage. Increasing the voltage swing to 2.4V (2 times higher) improves contrast ratio about 2 times (from 33% to 60%).

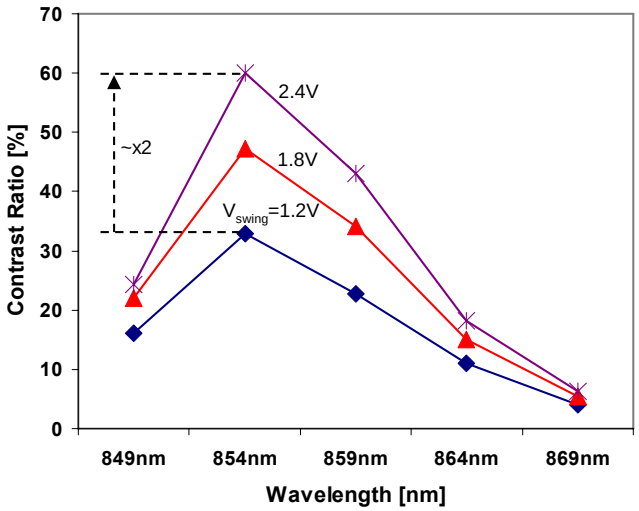
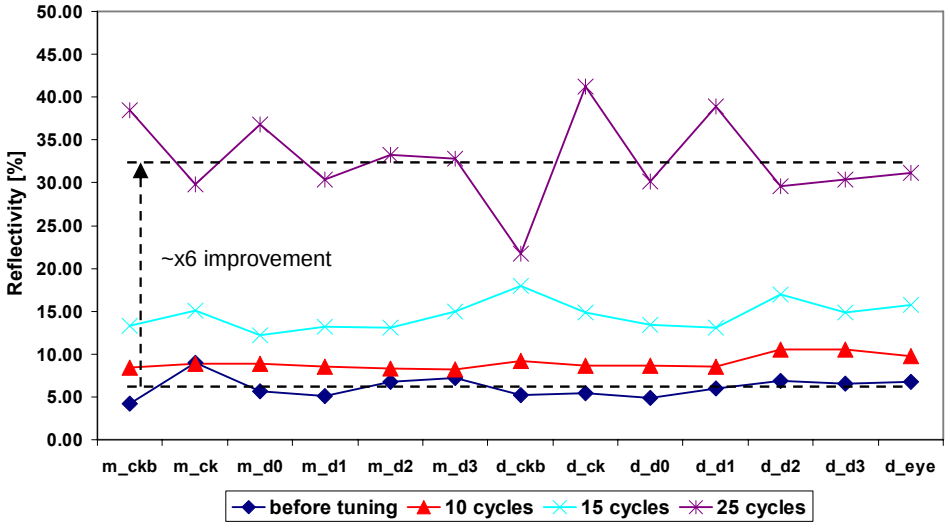
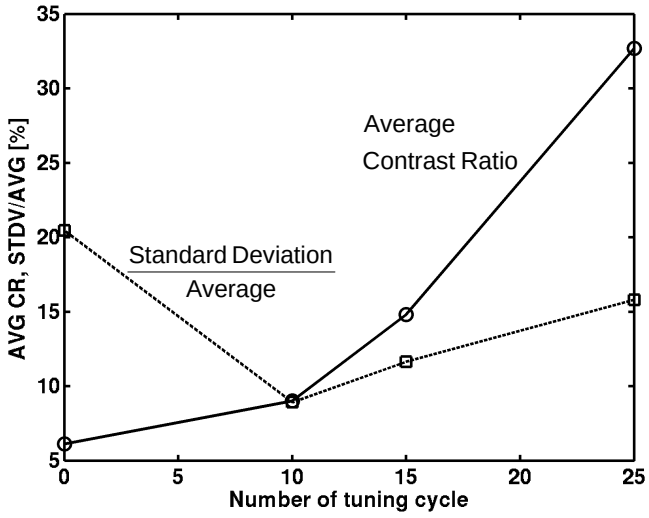


Fig.5.19 Contrast ratio vs. wavelength with variable swing voltages

The contrast ratio for different devices in a single chip is shown in Fig.5.20 as a function of the tuning cycle. The performance improved about 6 times (from 6% to 33%). We also quantify the variation of the contrast ratio using a metric dividing the standard deviation by the average that is shown in Fig.5.20(b). After the first couple of tuning cycles, the metric decreases quite a bit (less than 10%) and slightly increases with added tuning cycles. But the slope of the performance is pretty steeper than that of the metric, indicating the tuning process helps the performance while sacrificing the variation of the devices slightly.



(a)



(b)

Fig.5.20 (a) Improvement of the contrast ratio with tuning cycle for different devices (b) performance (contrast ratio) distribution in a single chip

5.5 Testing Result We Wish We Had

We designed the prototype electrical and optical interconnect in a single chip for verification of our models of power dissipation for off-chip communications. Unfortunately, the process variation shifts the internally generated DC bias controlling the current driving capability of the electrical transmitter, one of major components for testing the chip. This voltage shift results in insufficient driver capabilities to drive the network cable, hence, we could not get the signals, indicating the pass and fail of the communication transceivers. The engineering challenges in building such a system are nontrivial. For example, the concept of implementing both electrical and optical transceivers on a single chip requires overcoming many obstacles, for example, special package for optical interconnect, chip on board (COB), compromising the performance of electrical interconnect. The cavity tuning work with the specifically handled processing would be very useful for improving the performance of end-device for optical links. With our design techniques describe in this chapter and fixing the sensitivity issue of internal DC bias generating circuit, we could validate our model.

5.6 Summary

We present the circuit design of the prototype optoelectronic chip, implementing optical and electrical links in a single chip for more fair comparison. Source synchronous signaling is used for the clock-data recovery in both links. For electrical links, bipolar current mode simultaneous bidirectional signaling is implemented, maximizing bit rate for a pin-limited system. For optical link, on the other hand, transimpedance receiver with adjustable design parameters is used to optimize the receiver design, minimizing total optical link power. Offset tolerable

post-amplifier is followed to eliminate the process variation on the threshold of the inverter gain stage. We also discuss about the modulator and photo-detector device, and the fabrication processes to build the optoelectronic chip using a flip-chip bonding technique. By processing cavity tuning cycle, we improved the contrast ratio about 6 times (from 6% to 33%) with slightly sacrificing the variation in a single chip.

5.7 References

[1] W. J. Dally and J. H. Poulton, *Digital System Engineering*, Cambridge University Press, NY, 1998.

[2] S. Sidiropoulos and M. A. Horowitz, "A Semidigital Dual Delay-Locked Loop," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 11, pp. 1683-1692, Nov. 1997.

[3] P. Heydari and R. Mohavavelu, "Design of Ultra High-Speed CMOS CML Buffers and Latches," *Circuit and System*, vol. 2, pp. 208-211, May 2003.

[4] R. Rarjad-Rad, C. K. Yang, M. Horowitz, and T. Lee, "A 0.4-μm CMOS 10Gb/s 4PAM Serial Link Pre-emphasis Transmitter," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 580-585, May 1999.

[5] K. Lam, L. Dennison, and W. Dally, "Simultaneous Bi-directional Signaling for IC Systems," *Proceedings of the 1990 Conf. on Computer Design*, pp. 430-433, Sept. 1990.

[6] R. Mooney, C. Dike, and S. Borkar, "A 900Mb/s Bidirectional Signaling Scheme," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 12, pp. 2015-2023, Dec. 1995.

[7] K. Kim, S. Kim, G. Ahn, and D. Jeong, "A CMOS Serial Link for Fully Duplexed Data Communication," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 4, pp. 353-364, Apr. 1995.

[8] G. Ahn, D. K. Jeong, and G. Kim, "A 2-Gbaud 0.7-V Swing Voltage-Mode Driver and On-Chip Terminator for High-Speed NRZ Data Transmission," *IEEE Journal of Solid-State Circuits*, vol. 35, no.6, pp.915-918, June 2000.

[9] P. Mitran, F. Beaudoin, and M. N. El-Gamel, "A 2.5Gb/s CMOS Optical Receiver Frontend," *IEEE International Symposium Circuit and System*, vol.5, pp. V-441-V-443, 2002.

[10] M. Ingels, and M. S. J. Steyaert, "A 1Gb/s, 0.7μm CMOS Optical Receiver with Full Rail-to-Rail Output Swing," *IEEE Journal of Solid-States Circuits*, vol. 35, no. 7, pp. 971-977, July 1999.

[11] A. Cho, *Molecular Beam Epitaxy*, American Institute of Physics, Woodbury, 1994.

[12] G. A. Keeler, "Optical Interconnects to Silicon CMOS: Integrated Optoelectronic Modulators and Short Pulse Systems," Thesis of Stanford University, 2002.

[13] Y. H. Kuo, Y. K. Lee, Y. Ge, S. Ren, J. E. Roth, T. I. Kamins, D. A. B. Miller, and J. S. Harris, "Strong quantum-confined Stark effect in germanium quantum-well structures on silicon," *Nature*, vol. 437, no. 27, pp. 1334-1336, Oct. 2005.

[14] K. Giboney, J Simon, L. Mirkarimi, B Law, G Flower, S Corzine, M Leary, A Tandon, C Kocot, S Rana, A Grot, K. J. Lee, L Buckman, D Dolfi, "Next-generation parallel-optical data links," *IEEE Laser and Electro-Optics Society Annual Meeting 2001*, pp. 859-860, 2001.

[15] K. W. Goossen, J. A. Walker, L. A. D'Asaro, S. P. Hui, B. Tseng, R. Leibenguth, D. Kossives, D. D. Bacon, D. Dahringer, L. M. F. Chirovsky, A.L. Lentine, D. A. B. Miller, "GaAs MQW modulators integrated with silicon CMOS," *IEEE Photonic Technology Letter*, vol. 7, no. 4, pp. 360-362, 1995.

Chapter 6

Performance Comparison: On-chip

6.1 Introduction

Continuing current scaling paradigm of Cu/low-K based on-chip interconnects for global signaling presents a serious power and performance (latency) bottleneck in high-performance integrated circuits (IC). This is further exacerbated by the increase in copper resistivity, as wire dimensions and grain size become comparable to the bulk mean free path of electrons in copper (~40nm) [1]. On the architecture side, the shift toward a multi-core paradigm would demand a high bandwidth/bandwidth density and low latency connections between cores. Bandwidth density is defined as bandwidth per unit distance. Thus, it is imperative to examine alternate interconnect schemes for future ICs. The two most important novel potential candidates are optical and carbon nanotube (CNT)-based interconnects.

Optical interconnects can potentially reduce latency and provide high-bandwidth with relatively low power dissipation [2][3]. However, the medium for an

optical interconnect (a waveguide), has a relatively larger size (pitch~0.6μm), making it difficult to provide high bandwidth density. This problem can be mitigated using wavelength division multiplexing (WDM) with 10 to 40 wavelengths. CNT interconnects, on the other hand, have the flexibility of being implemented in the same size scale as the existing Cu wires (~50nm), hence, can provide high bandwidth density. Further, single-wall CNT (SWCNT) can yield a much lower latency compared to Cu/low-k by virtue of having a much higher mean free path. The high quality SWCNT mean free path can be in the micron range [4].

In this chapter, we compare CNT-based and optical interconnects with Cu interconnects using commonly deployed metrics, latency and power. However, latency and power by themselves are not necessarily the most complete metrics for comparing different interconnects, as a larger bandwidth and a smaller latency can be obtained using more area resources. To account for this issue, Naeemi et al. [5] proposed bandwidth density/latency as an appropriate comparison metric. We augment this metric by further normalizing it with respect to power dissipation to account for the fact that power can also be used to increase bandwidth. We assume that the serial bit rate per link is limited by the global clock frequency (FO4 inverter delay) which is not expected to increase dramatically in the future. We divide the bandwidth by the pitch, the latency, and the power to obtain the bandwidth density/latency/power metric. This metric is important because for global communication between blocks/cores, the designers care about the bandwidth density, the latency of the links and how much power they need to expend to obtain the required link speeds. We extensively examine how this metric is affected by 1) the device parameters – the modulator and the detector capacitances, in the case of optical interconnects, 2) the materials parameters - the mean free path and the packing density, for CNTs, and 3) the system parameters - the global clock frequency and the switching activity.

The rest of the chapter is organized as follows: Section 6.2 describes how the circuit model parameters-resistance, capacitance, and inductance for Cu interconnects were obtained. Section 6.3, does the same for a SWCNT and a CNT bundle. Using these circuit models, section 6.4 demonstrates the repeater optimization for Cu and SWCNT and obtains latency and power. Section 6.5 describes the performance modeling for the quantum-well modulator-based optical interconnect and its latency and power. In section 6.6, we compare a system's driven, top-down, compound metric for Cu, CNTs, and optics and comprehensively examine the impact of device, material, and system parameters on it. Finally, we conclude in section 6.7.

6.2 Cu Circuit Parameter Modeling

As Cu wire dimensions and grain size become comparable to the bulk mean free path of electrons (~40nm at room temperature), electrons experience surface and grain-boundary scattering, which, in turn, increases its electrical resistivity. The Fuchas and Sondheimer (F-S) model is used for surface scattering effect [6],

$$\frac{\rho}{\rho_o} = 1 + \frac{3}{4}(1-p)\frac{l}{w} \tag{6.1}$$

where, ρ_o is the resistivity of the bulk (1.9μΩ·cm), p is the fraction of electrons scattered specularly at the surface (assumed 0.6 [1]), w is the width of the wire (taken from ITRS [7]), and l is the mean free path of the bulk material. In order to model the scattering at grain boundaries, the theory of Mayadas and Shatzkes (M-S model) was applied [8].

$$\frac{\rho}{\rho_o} = 3\left[\frac{1}{3} - \frac{\alpha}{2} + \alpha^2 - \alpha^3 \ln(1+\frac{1}{\alpha})\right]$$
$$\alpha = \frac{l}{d}\frac{R}{1-R} \tag{6.2}$$

where, R is the reflectivity coefficient describing the fraction of electrons that are not scattered by the potential barrier at a grain boundary (assumed 0.5 [1]) and d is the average distance of grain boundaries (assumed wire width; d=w). Fig.6.1 shows Cu resistivity in terms of wire width. For 22nm technology node, resistivity increases to 5.8 μΩ·cm, about 3X of the bulk resistivity. This dramatically increased resistivity adversely impacts the performance of Cu wires.

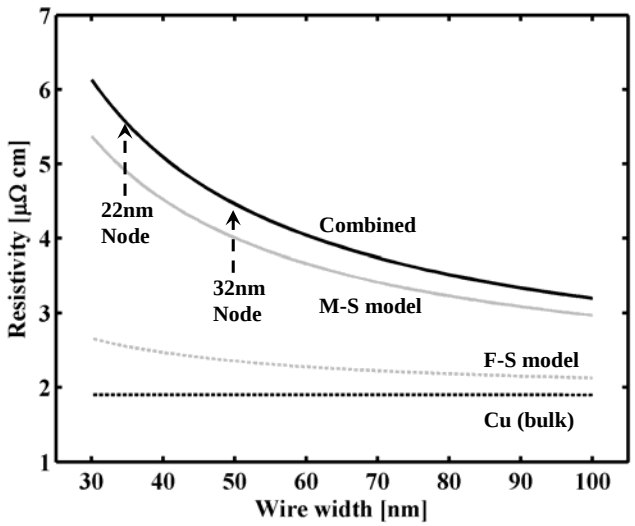


Fig.6.1 Cu resistivity as a function of wire width taking into account of both surface and grain boundary scattering. p and R values were assumed to be 0.6 and 0.5, respectively.

The wire capacitance is evaluated using an analytic model [9],

$$C_w = \varepsilon \left[\begin{array}{l} 1.15\frac{w}{t} + 2.80\left(\frac{h}{t}\right)^{0.222} \\ + \left(0.66\frac{w}{t} + 1.66\frac{h}{t} - 0.14\left(\frac{h}{t}\right)^{0.222} \right) \cdot \left(\frac{t}{s}\right)^{1.34} \end{array} \right] \quad (6.3)$$

where, ε is the permittivity of dielectric between a wire and a ground plane, s is the space between wires (assumed $s=w$), h is the height of the wire ($h=w\times$aspect ratio), and t is the distance between metal layers (assumed $t=4h$).

In this work, we also included the impact of inductance. Its value for a wire with a rectangular cross-section area can be expressed as following [10].

$$L_w = 2\times10^{-7} l\left(\ln\frac{2l}{w+h} + 0.5 + \frac{w+h}{3l} \right) \quad (6.4)$$

This comes out to be about 0.5nH/mm at 22nm technology node.

6.3 CNT Circuit Parameter Modeling

6.3.1 Single Wall CNT Model

A SWCNT has two propagating channels because of its band structure (sublattice degeneracy). In addition, electrons can have spin up or spin down. The equivalent circuit model is shown schematically in Fig.6.2, and its components will be explained in the following subsections [11].

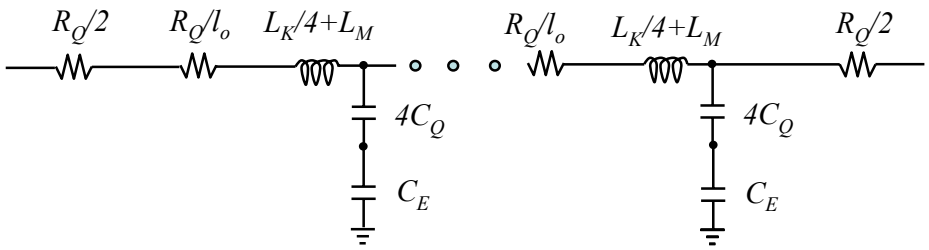


Fig.6.2 Equivalent circuit model of a SWCNT interconnect. Resistance (*R*), inductance (*L*), and capacitance (*C*) are values per unit length.

6.3.2 Inductances

In the presence of a ground plane, the magnetic inductance per unit length is given by [12]

$$L_M = \frac{\mu}{2\pi}\cosh^{-1}\left(\frac{2t}{d_t}\right) \approx \frac{\mu}{2\pi}\ln\left(\frac{t}{d_t}\right) \tag{6.5}$$

where, μ is the permeability (=$4\pi\times10^{-7}$) and d_t is the nanotube diameter. For d_t =1nm and $t=4h$=340nm, magnetic inductance is about 1.2nH/mm. In addition to the magnetic inductance, because of a low density of state in the conduction band of SWCNT, the kinetic energy stored in the current carriers (electrons) is very large. This

results in a very high kinetic inductance per unit length given by [12]

$$L_K = \frac{h}{2e^2 v_f} \tag{6.6}$$

where, h is Plank constant, v_f is the Fermi velocity of electrons in carbon nanotube (8×10^5m/s). The kinetic inductance is about 16μH/mm, which is about four orders of magnitude larger than its magnetic inductance per CNT.

6.3.3 Capacitances

The electrostatic capacitance is given by [12]

$$C_E = \frac{2\pi\varepsilon}{\cosh^{-1}(2t/d_t)} \approx \frac{2\pi\varepsilon}{\ln(t/d_t)} \tag{6.7}$$

where, ε is the permittivity of dielectric between a wire and a ground plane. For d_t=1nm, t=*4h*=340nm, and ε_r (relative permittivity for 22nm technology node)=2.0, the electrostatic capacitance is about 190fF/mm. In addition, there is a non-negligible quantum capacitance associated with a CNT. To add an electron in a SWCNT, one must add it at an available quantum state above the Fermi energy (E_F) due to the Pauli's exclusion principle. The quantum capacitance per unit length for a CNT is given by [11]

$$C_Q = \frac{2e^2}{hv_f} \tag{6.8}$$

This capacitance is around 100fF/mm, which is of the same order of magnitude as its electrostatic counterpart.

6.3.4 Resistance

The resistivity of a conducting material depends on the number of scattering events, the number of carriers, and the availability of the states. If there is no scattering at the contact, using the Landauer-Buttker formula, the fundamental resistance (quantum resistance, R_Q) is about 6.45KΩ. As shown in Fig.6.2 this quantum resistance is equally divided between the two contacts on either side of the nanotube.

Electrons moving along the CNT can get scattered by either defects or by phonons, hence, possess a finite mean free path. Both linear and exponential dependencies of CNT resistance on length have been reported. However, high quality CNTs having a large mean free path exhibit a linear dependence. We assume a linear dependence. This has also been corroborated by recent experiments [13][14]. Thus,

$$R_{SWCNT} = R_Q\left(1+\frac{l}{l_o}\right) \tag{6.9}$$

where, l is wire length, l_o is the mean free path of electrons.

The electron mean free path of SWCNT shows a linear relation with the SWCNT diameter.

$$l_o = C_l \cdot d_t \tag{6.10}$$

The proportionality constant (C_l) is 2.8μm/nm using theoretical calculations [4] (henceforth, ideal model) and 0.9μm/nm using experimental fitting (henceforth, practical model) [13]. We assumed SWCNT diameter of 1nm, thus ideal model yields a mean free path of 2.8μm, whereas the practical model results in a value of 0.9μm. The aforementioned mean free paths are valid at smaller bias voltages. At higher voltages (above the critical bias=0.16V), optical and zone-boundary phonon scatter the electrons, hence the mean free path decreases to approximately 30nm [13]. A rigorous analysis shows that for global wires the voltages across the wires is always less than the critical bias [15] lending credibility to mean free path assumption of the order of μm.

6.3.5 CNT Bundle

The circuit model parameters for a CNT bundle is calculated by treating the individual tubes in parallel. Hence, it depends on the total wire cross sectional area. The packing density (*PD*) of a CNT bundle, dictated by the fraction of SWCNT exhibiting a metallic behavior, is taken into account by considering the space between SWCNTs (x) [15] as shown in Fig.6.3.

The number of SWCNTs in a bundle is given by [16],

$$
\begin{aligned}
n_w &= \frac{w - d_t}{x}, n_h = \frac{h - d_t}{\left(\sqrt{3}/2\right)x} + 1 \\
n_{CNT} &= n_w n_h - n_h / 2 \qquad \text{if } n_h \text{ is even} \\
&= n_w n_h - \frac{n_h - 1}{2} \qquad \text{if } n_h \text{ is odd}
\end{aligned}
\tag{6.11}
$$

where, n_w is the number of "columns" in a bundle, n_h is the number of "rows" in a bundle, and n_{CNT} is the number of SWCNTs in a bundle.

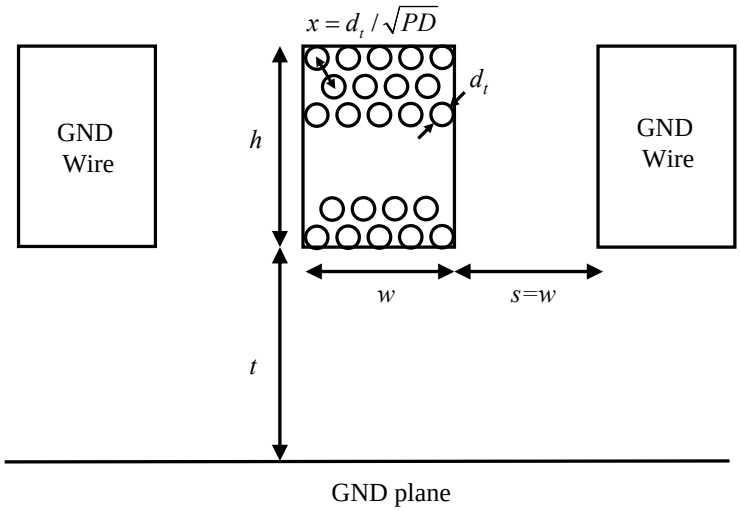


Fig.6.3 Schematic of the interconnect geometry with CNT bundles, taking packing density (*PD*) into account.

The resistance of a CNT bundle is simply the quantum resistance divided by n_{CNT}.

$$R_w = \frac{R_Q}{n_{CNT}}\left(1+\frac{l}{l_o}\right) \tag{6.12}$$

Scattering at the contacts results in contact resistance that may be up to hundreds KΩ for poor contacts. However, there are many recent reports indicating that the contact resistance can be brought down to a few KΩ [17][18] per tube. In a bundle, this value reduces substantially and can be ignored compared to the wire resistance of long global wires.

The effective quantum capacitance was found to be negligible compared to its electrostatic counterpart as it is substantially reduced in a bundle. On the other hand, there have been conflicting reports in literature regarding the magnitude of the electrostatic capacitance of CNTs. Reference [16] claims two times higher magnitude compared to Cu wires of similar dimensions, based on the surface curvature of CNTs (~1nm roughness). However, using FastCap (3D field solver), we find that the surface curvature of CNT only contributes to about 4% increase in electrostatic capacitance. Our results are in agreement with the work in reference [19]. Hence, we assumed the capacitance of a CNT bundle is the same as the electrostatic capacitance of Cu wire.

The kinetic inductance of a CNT bundle is given by the parallel combination of the kinetic inductances of each SWCNT. A value of ~6nH/mm was obtained for wire dimensions corresponding to the 22nm technology node and with a CNT diameter of 1nm. In contrast, the magnetic inductance remains relatively constant with wire dimensions at around ~1.6nH/mm. We account for the mutual inductance between SWCNTs in a bundle [20] using partial element equivalent circuit (PEEC) model [21]. The total inductance is given by

$$L_w = \left(\frac{L_K}{4n_{CNT}} + L_M \right) \cdot l \tag{6.13}$$

Since inductance is important and the resistance is lower than Cu wires, it is imperative to use RLC, as apposed to RC, circuit models.

6.4 Repeater Optimization for Cu and CNT Interconnects

In evaluating the performance metrics of the global/semiglobal wires we use an RLC circuit model described in the previous sections and include repeater insertion for delay reduction. Fig.6.4 shows the equivalent circuit model for the CNT interconnects with the repeaters, including the kinetic inductance and the contact resistance (negligible for global wires). In contrast with an RC wire model, for an RLC wire, there are no closed form solutions optimizing the repeater size (k) and its spacing (h). We optimize these parameters (for minimum wire delay) using Newton-Raphson numerical iteration method along the lines of the methodology outlined in the reference [22]. The increase in the ratio of the inductance to the resistance results in a smaller optimum repeater size and a larger optimum repeater spacing. Thus, the total repeater capacitance reduces resulting in a lower power dissipation.

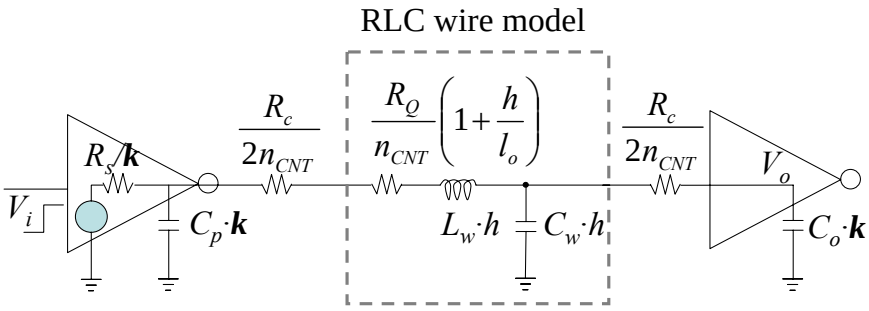


Fig.6.4 Equivalent circuit model of a repeater segment for CNTs.

Fig.6.5 shows the power and the latency in terms of the bandwidth density for Cu and a CNT bundle with different global clock frequencies at the 22nm technology node (henceforth, we assumed 22nm technology node). At a given global clock frequency, the different bandwidth densities are achieved by implicitly varying the wire width (pitch). The reason for having the bandwidth density as an independent variable in the subsequent plots is because we expect this to be the required metric set by the designers using the system/architectural constrains, especially between multiple cores in a high performance IC. The maximum achievable bandwidth density is limited by the minimum wire width (from ITRS [7]) as shown in Fig.6.5.

As wire pitch increases (lower bandwidth density), at first, power decreases, but at very low bandwidth density it starts to go back up. This is a direct result of power being proportional to the wire and repeater capacitance (CV^2f). Capacitance has two components: inter-metal, C_{IMD}, and interlevel, C_{ILD}. For a given thickness of the metal, these components follow opposite trends (inset in Fig.6.5), yielding a minimum in the total capacitance with respect to the wire pitch. In addition, the capacitance of the optimized repeaters tracks wire capacitance and is a fraction of it. From Fig.6.5 we also observe that the CNT bundle has a lower power dissipation than Cu wires for larger wire pitch. This is because a lower resistance in this interconnect prompts an RLC, rather than an RC behavior, which results in larger repeater spacing and a

consequent capacitance reduction. In addition, Fig.6.5 shows that the latency for both Cu and CNT bundle increases with bandwidth density. This is because larger bandwidth density (lower wire width) results in a larger wire resistance, which in turn results in more repeaters per wire. Furthermore, CNTs are shown to exhibit about 1.7X lower latency compared to Cu/Low-K systems owing to their dramatically lower resistance.

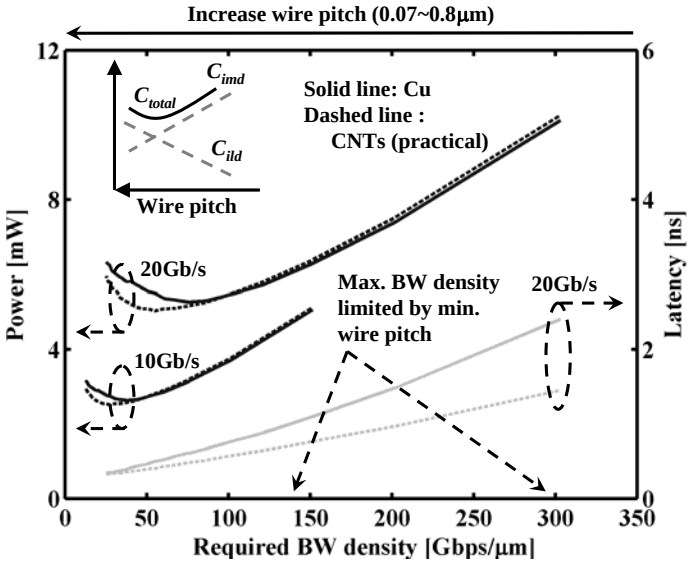


Fig.6.5 Power and latency of Cu and CNTs (practical) as a function of required bandwidth (BW) density for 10mm wire length with different global clock frequencies. The implicit parameter being varied is the wire pitch to obtain varying BW density. We assumed 1/3 packing density (*PD*) for CNTs and 100% switching activity for both Cu and CNTs.

6.5 Circuit Model for Optical Interconnects

The latency and power analysis of optical interconnect was based on a system consisting of quantum-well modulator at the transmitter, waveguides for transmission medium, and a transimpedance amplifier (TIA) at the receiver (Fig.2.1). The transmitter latency is given by the product of the technology dependent, fan-out-four (FO4) inverter delay, and the number of buffers driving the modulator capacitance. To calculate the receiver latency, we assumed that the input pole (the node of detector capacitance and input of TIR) is dominant. Finally, the delay of the waveguide is approximately dictated by the speed of light in the waveguide core material (silicon) (11.3ps/mm). The total power dissipation for the optical interconnect is calculated by optimizing the sum of the receiver and the transmitter powers. The receiver power is based on its design parameters (front-end transistor size, feedback resistance (R_f), number of post-amplifier stage) as outlined in Appendix A.

Fig.6.6 shows the power and latency of optical interconnect as a function of bandwidth density. This plot is analogous to the electrical plots for Cu and CNT in Fig.6.5. However, the implicit parameter, which changes the bandwidth density, is the number of multiplexed channels as apposed to the wire pitch (Cu/CNT interconnects). For optical interconnects, the wire (waveguide) width is relatively fixed at around 0.3μm, constrained by the large radiation losses below this value. However, wavelength division multiplexing (WDM) can be used to achieve the same purpose. Now, the maximum bandwidth density is limited by the total number of wavelengths that can be multiplexed. Without WDM, the power follows a relatively linear relation with the bandwidth at lower values, however, it starts to increase super-linearly beyond a certain value depending on the maximum available laser power. The performance of the optical interconnects is highly dependent on the detector and the

modulator capacitances (C_{det}, C_{mod}). Two different values of 10 and 50fF are shown, representing monolithically integrated detector and a hybrid III-V bonded detector, respectively. In addition, Fig.6.6 depicts the latency as a function of the bandwidth density. Since it is not a function of the number multiplexed channel, it shows up as a straight line. For higher global clock frequency, to satisfy the required bandwidth, the effective transimpedance of the TIR is reduced, hence, it reduces the latency, while, the input buffer size increases and it results in a larger power dissipation.

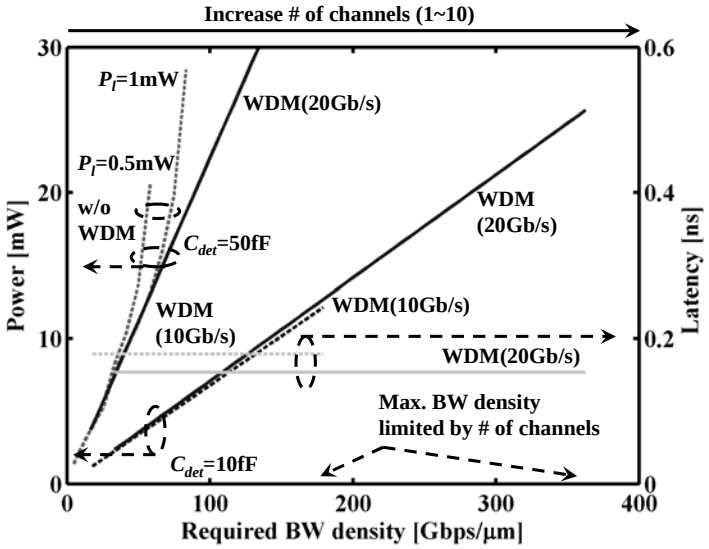


Fig.6.6 Power and latency of optical interconnects as a function of bandwidth (BW) density using WDM (up to 10 channels) at 10mm wire length with different global clock frequencies. We evaluated for two device capacitances (C_{det} =C_{mod} =10fF and 50fF) assuming coupling loss of 3dB, waveguide loss of 0.2dB/cm, and 100% switching activity.

6.6 Performance Comparison

We now compare the various performance metrics including the power density, the latency, as well as the compound performance metric-bandwidth density/latency /power. Further, we evaluate the impact of the device, the material and the system parameters on the compound metric. Hereafter, the wire length is assumed to be 10mm corresponding to the global wires.

6.6.1 Latency and Power

Fig.6.7 compares the power density and the latency of the three interconnects in terms of the bandwidth density. Given the core/block dimensions, this plot can be used to calculate the total power expended for the required aggregate bandwidth between cores/blocks. Several interesting observations can be made: (i) The latency of the optical interconnects is much lower than both the Cu and the CNT bundle. At high required bandwidth densities this advantage is about 13X and 8X compared to the Cu and the CNT bundle, respectively. The advantage diminishes at lower bandwidth densities. A CNT bundle exhibits ~ 1.7X lower latency than the Cu interconnect at high bandwidth densities. (ii) The optical interconnect consumes lower power than both the Cu and the CNT interconnect. It indicates optics is more power-efficient for the same required aggregate bandwidth between cores. (iii) Optical interconnects allow a much higher absolute bandwidth density than the Cu or the CNT bundle because of the WDM option. In addition, the maximum bandwidth density for Cu and a CNT bundle may be lower than shown here because an excessive number of repeaters may present a more stringent practical limit on the minimum attainable wire pitch.

Fig.6.7 also aids in contrasting the fundamental differences between electrical (Cu and CNTs) and optical interconnects, borne out of the stark differences in the dependencies (or lack of them) between power/latency and the bandwidth density. For example, the latency is independent of the bandwidth density for optical interconnects, whereas it increases with the bandwidth density in the case of electrical interconnects. Also, optical interconnects use multiplexing and not wire pitch shrinkage, implicitly, to achieve higher bandwidth density. These differences point to the importance of architectural differences to maximally exploit the potential of each type of interconnect.

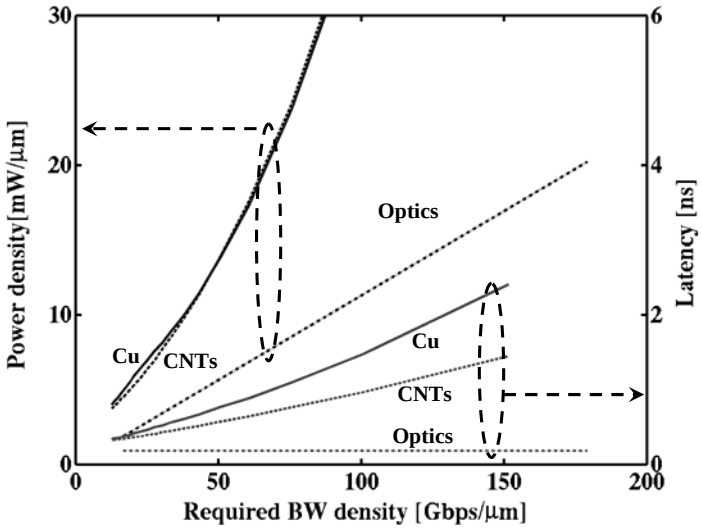


Fig.6.7 Power density and latency comparison bwtween Cu, CNTs (practical), and optics as a function of bandwidth (BW) density at 10Gb/s global clock frequecny and 10mm wire length, and 100% switching activity.

6.6.2 Compound Metrics: Bandwidth Density per Delay per Power

The normalization of the bandwidth density with power is imperative because power can be used as a knob to increase the bandwidth. In addition, a normalization with the latency is important as it can contribute to the communication bottleneck depending on the nature of the data traffic. Thus, bandwidth density per latency per power severs as a fair compound metric to compare various interconnect technologies. For this metric, bandwidth density per latency is the measure of the interconnect performance and the power is the price to attain that performance.

Fig.6.8 compares the bandwidth density per power per latency as a function of bandwidth density for all interconnects under consideration. For optical interconnects, this performance metric is constant as the increase in the number of multiplexed channels linearly increases the power and does not change the latency. However, for the Cu and CNT technology, the compound metric increases at lower bandwidth density, and decreases at the higher values because of various competing factors: At low values, as bandwidth density increases, power dissipation reduces, but is compensated by an increase in the latency (Fig.6.5). However, at higher bandwidth density both the power and the latency rise with increase in the bandwidth density, lowering the compound metric (Fig.6.8). The inset in Fig.6.8 shows the trends for each of the individual components constituting the compound metric. There exists an optimum bandwidth density (and a corresponding wire pitch), which maximizes this metric. It occurs at a wire pitch of about 70~100nm. Even at this optimum point, corresponding to the best case for Cu and CNT wires, the optical interconnects outperform the CNTs by 2.5X and the Cu wires by 3.4X in terms of this metric. In this figure, we use a constant global clock frequency based on a trend toward slowdown of the clock frequency and a paradigm shift toward multiple cores. If the global clock frequency was increased, it will extend the maximum achievable bandwidth density

for all interconnects, however, it will not improve the performance metric for the CNTs and the Cu wire because of the normalization. However, the metric will improve slightly for the optical interconnects (13%) due to the reduction in the latency as shown in Fig.6.6.

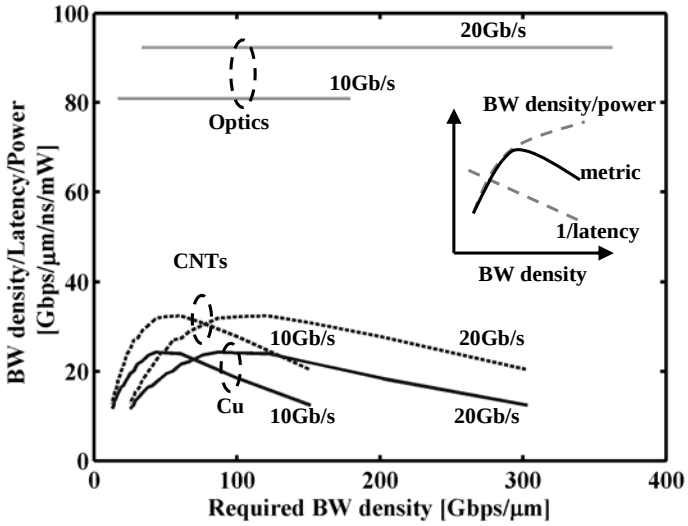


Fig.6.8 Compound metric comparison between Cu, CNTs (practical), and optics at two different global clock frequencies (10Gb/s and 20Gb/s), 10mm wire length, and 100% switching activity.

The performance of the CNTs is a strong function of the electron mean free path and the packing density. Fig.6.9 shows the effect of these CNT parameters on the compound performance metric. As the packing density improved from 1/3 (practical) to 1 (ideal), CNTs close the gap with optical interconnects from 2.6X to 1.3X. Similarly, the improvement in the mean free path from 0.9μm (practical) to 2.8μm (ideal) also decreases the optical interconnect advantage over CNTs to about 1.2X. With both ideal mean free path and packing density, the CNTs can outperform the optical interconnects by 40% in terms of this metric. In this figure C_{det}=C_{mod}=10fF

was chosen for optical interconnects and can be achieved using a monolithically integrated Ge Metal-Semiconductor-Metal photo-detector [24] and SiGe quantum-well modulator [24]. However, because the device capacitances can be further decreased, optical interconnects can potentially be improved.

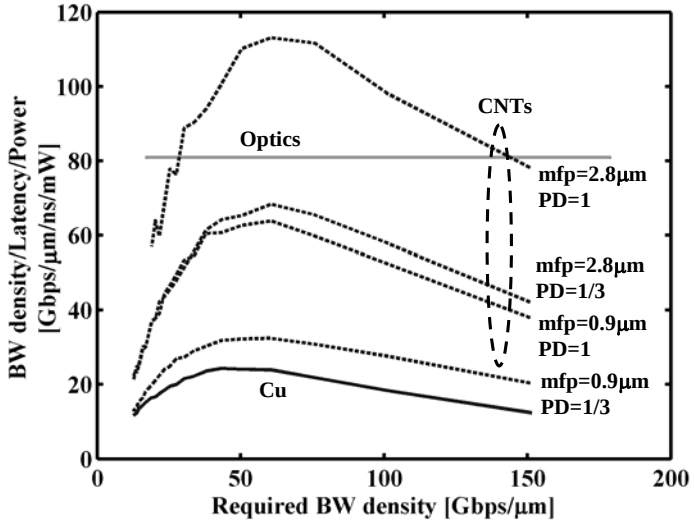


Fig.6.9 Compound metric comparison between Cu, CNTs, and optics with different mean free path and packing density (*PD*) for CNTs at 10Gb/s global clock frequency, 10mm wire length and 100% switching activity.

A fundamental difference between the electrical and the optical interconnects is in the nature of the power dissipation. A large portion of the power in the optical system is the static power, which is independent of the switching activity. While, electrical wires are dominated by the dynamic power, which is linearly dependent on the switching activity. Thus, it is imperative to examine the comparisons in terms of the switching activity, in case future wire architectures using wires more efficiently are designed. Fig.6.10 shows the maximum value of the compound performance metric for different switching activities. For 50fF device capacitances (current flip-chip bonded detector technology), optics is between Cu and CNTs (both practical and

ideal) at high switching activity. However, less than 70% switching activity, Cu can have better performance compared to optics. However, with 10fF device capacitances, optics is comparable in performance up to 25% and 35% switching activity compared to Cu and CNTs (practical), respectively.

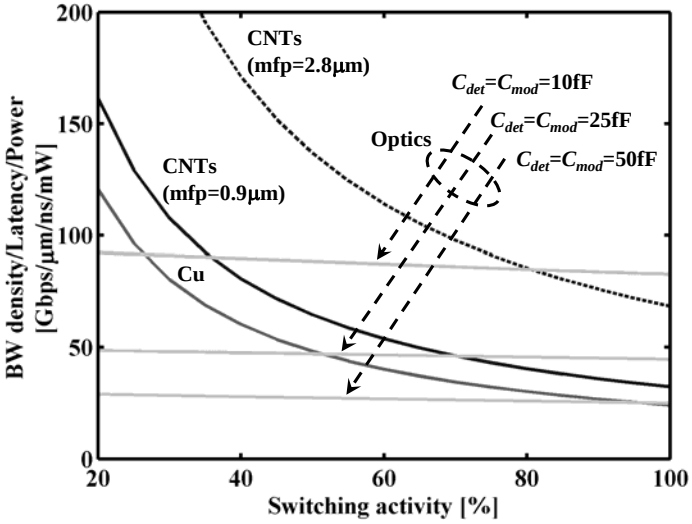


Fig.6.10 Compound metric comparison in terms of switching activity between Cu, CNTs, and optics.

6.7 Summary

We have quantified the accurate circuit model (*R*, *L*, and *C*) for CNTs, incorporating the mutual inductance between SWCNTs in a bundle. Using these models, we have compared the latency and power of Cu, CNTs, and optical interconnects. Optical interconnects have the lowest latency for global levels, whereas Cu and CNTs have lower power at the same bandwidth density because of large wire pitch (~0.6μm) of optical interconnects. We also compared the compound performance metric, bandwidth density/latency/power, for a fair comparison as a

system driven top-down metric. We find that the Cu and CNT interconnects posses an optimum wire width maximizing this metric at about 70~100nm range. Optical interconnects (C_{det}=C_{mod}=10fF) have 2.5X and 3.4X better performance metric than CNTs (practical) and Cu, respectively. With an improvement in the mean free path (~2.8μm) or the packing density (100%) for CNTs, the performance advantage for optics can be reduced to 30% and 20%, respectively. Finally, the optical interconnects yield better performance beyond 35% and 20% switching activity compared to the CNTs (practical) and the Cu wires, respectively.

6.8 References

[1] W. Steinhogl, G. Schindler, and M. Engelhardt, "Size-dependent resistivity of metallic wires in the mesoscopic range," *Physics Review B*, vol. 66, 075414, 2002.

[2] H. Cho, P. Kapur, and K. C. Saraswat, "Power Comparison between High-speed Electrical and Optical Interconnects for Inter-chip Communication," *Journal of Lightwave Technology*, vol. 22, no. 9, pp. 2021-2033, 2004.

[3] P. Kapur, and K. C. Saraswat, "Comparisons between electrical and optical interconnects for on-chip signaling," *International Interconnect Technology Conference*, pp. 89-91, 2002.

[4] C. T. White, and T. N. Todorov, "Carbon Nanotube as Long Ballistic Conductors," *Nature*, vol. 393, pp. 240-242, 1998.

[5] A. Naeemi, R. Venkatesan, and J. D. Meindl, "Optimal Global Interconnects for GSI," *Transactions on Electron Devices*, vol. 50, no. 4, 2003.

[6] E. H. Sondheimer, "The mean free path of electrons in metals," *Advanced Physics*, vol. 1, no. 1, pp.1-2, 1952.

[7] International technology roadmap of semiconductor, 2004 version

[8] A. F. Mayadas and M. Shatzkes, "Electrical-Resistivity Model for Polycrystalline Films: the Case of Arbitrary Reflection at External Surfaces," *Physics Review B*, vol. 1, pp. 1382-1389, 1970.

[9] T. Sakurai and T. Tamuru, "Simple Formulas for two- and three-dimensional capacitances," *Transactions on Electron Devices*, vol. 30, pp. 183-185, 1983.

[10] C. P. Yue, and S. S. Wong, "Physical Modeling of Spiral Inductors on Silicon," *Transactions on Electron Device*, vol. 47, no. 3, pp. 560-568, 2000.

[11] P. J. Burke, "Luttinger Liquid Theory as a Model of the Gigahertz Electrical Properties of Carbon Nanotubes," *Transactions on Nanotechnology*, vol. 1, no. 3, pp. 129-144, 2002.

[12] S. Ramo, J. R. Whinnery, and T. V. Duzer, *Field and Waves in Communication Electronics*, New York: Wiley, 1994.

[13] J. Y. Park, S. Rosenbelt, Y. Yaish, V. Sazonova, H. Ustunel, S. Braig, T. A. Arias, P. L. McEuen, "Electron-phonon Scattering in Metallic Single-wall Carbon Nanotubes," *Nano Letters*, vol.4, pp. 517-520, 2004

[14] S. Li, Z. Yu, C. Rutherglen, and P. J. Burke, "Electrical Properties of 0.4cm long Single-walled Carbon Nanotube," *Nano Letters*, vol.4, pp.2003-2007, 2004.

[15] A. Naeemi, and J. D. Meindl, "Design and Performance Modeling for Single-Wall Carbon Nanotubes as Local, Semi-global and Global Interconnects in Gigascale Integrated Systems," to be appeared in *Transactions on Electron Device*.

[16] N. Srivastava, and K. Banerjee, "Performance Analysis of Carbon Nanotube Interconnects for VLSI Application," *IEEE/ACM International Conference on Computer-Aided Design*, pp. 383-390, 2005.

[17] Z. Yao, C. L. Kane, and C. Dekker, "High-field Electrical Transport in Single-wall Carbon Nanotubes," *Physics Review Letters*, vol. 84, pp. 2941-2944, 2000.

[18] O. Hjortstam, P. Isberg, S. Soderholm, and H. Dai,"Can We Achieve Ultra Low Resistivity in Carbon Nanotube-based metal Composites?" *Applied Physics A*, vol. 78, pp. 1175-1179, 2004.

[19] A. Naeemi, R. Sarvari, and J. D. Meindl, "Performance Comparison between Carbon Nanotube and Copper Interconnect for Gigascale Integration (GSI)," *Electron Device Letters*, vol. 26, no. 2, pp. 84-86, 2005.

[20] A. Nieuwoudt, and Y. Massoud, "Evaluating the Impact of Resistance in Carbon Nanotube Bundles for VLSI Interconnect Using Diameter-Dependent Modeling Technique," *Transactions on Electron Device*, vol. 53, no. 10, pp. 2460-2466, 2006.

[21] M. W. Beattie, and L. T. Pileggi, "Inductance 101: Modeling and Extraction," *Proceeding IEEE DAC*, 2001.

[22] K. Banerjee, and A. Mehrotra, “Analysis of On-chip Inductance Effects for Distributed RLC Interconnects,” *Transactions on Computer-Aided Design of Int. Circuits and Systems*, vol. 21, no. 8, pp. 904-914, 2002.

[23] O. Kibar, D. A. V. Blerkom, C. Fan, and S. C. Esener, “Power Minimization and Technology Comparisons for Digital Free-Space Optoelectronic Interconnects,” *Journal of Lightwave Technology*, vol. 17, no.4, pp. 546-554, Apr. 1999.

[24] D. Buca et. al., “Metal-Germanium-Metal ultrafast infrared detectors,” *Journal of Applied Physics*, vol. 92, no. 12, pp. 7599-7605, Dec. 2002.

[25] Y. H. Kuo, Y. K. Lee, Y. Ge, S. Ren, J. E. Roth, T. I. Kamins, D. A. B. Miller, and J. S. Harris, “Strong quantum-confined Stark effect in germanium quantum-well structures on silicon,” *Nature*, vol. 437, no. 27, pp. 1334-1336, Oct. 2005.

Chapter 7

Summary and Future Recommendations

7.1 Summary

In this dissertation, we have performed performance comparisons between Cu, CNT-based, and optical interconnects for on-chip and off-chip communications.

For off-chip applications, we compared high speed optical and electrical interconnects using relevant metrics, such as power and bandwidth. Part of the comparison entailed optimization of the link power for a given bandwidth. We find that for a given communication bandwidth requirement, beyond a critical length, power optimized optical interconnects dissipates lower power compared to high-speed electrical signaling schemes. Beyond the 32nm technology node, with its commensurate bandwidth requirement, optical interconnects becomes favorable for distances as little as 10cm. These distances correspond to inter-chip communication on a board. The critical distances are calculated based on reasonable assumptions for the electrical and optical end-device specifications including optical detector/modulator

capacitances, optical coupling efficiency, and electrical receiver's sensitivity and offsets. We also assess the impact of the improvement in these parameters on system's power/bandwidth performance and the critical length. The sensitivity analysis gives device designers a system's evaluation framework.

In addition, we examined two competing optical transmitter technologies for off-chip link: the vertical cavity surface emitting laser (VCSEL) and the quantum well modulator (QWM), for optical links. VCSEL-based links are power favorable compared to QWM-based links at higher bandwidth (>20Gb/s) and larger distance. However, from the practical standpoint, it is instructive to note that currently VCSELs are difficult to drive beyond 15-20Gbps and are like to suffer from reliability problems in harsh CMOS chip environment with elevated temperatures. The QWM is better for lower detector and modulator capacitances. Furthermore, we quantify the design constraints on the modulator under which it is superior to VCSEL technology as a function of bandwidth, link length, and transmitter and detector capacitances. We also identify the range of tolerable insertion loss in QWM. We find that to render QWM more power-efficient compared to VCSELs, lower capacitances (<15fF) for both the modulator and the photo-detector would be required. Recently, germanium (Ge) quantum wells with high germanium concentration, SiGe barriers, monolithically integrated on silicon substrates, are shown to exhibit strong quantum confined stark effect (QCSE). This can decrease the device capacitances below than 10fF [1].

For off-chip links we further related the design parameters of a QWM including pre-bias voltage and the number of quantum wells, and the required voltage swing, with the total link power. In this regard, we presented an optimization methodology for minimizing total optical link power, and obtained the optimized modulator design parameters (number of wells, pre-bias voltage) and system parameters (swing voltage), which accomplish this. As bit rate increases, the optimum

voltage swing exceeds the stipulated CMOS supply voltage at the 65nm technology node. An operation at the sub-optimal voltage, constrained by the maximum available voltage swing, results in about 46% power penalty at 25Gb/s compared to optimum voltage swing if that was available. A large power penalty for sub-optimum supply operation makes a strong case for either a different supply voltage for the I/O modulator driver (this is not inconceivable in the light of an already existing need to have more than one customized supply voltage on a CMOS chip for different functional blocks) or a charge pump to boost up the voltage swing.

We also examined the impact of the device and system parameters (laser power, coupling and propagation loss, bit rate, and capacitances) on both the optimum modulator design and operation parameters as well as the optimum link power. A higher bit rate results in an increase in the noise bandwidth and other noise sources, thus a larger rms noise at the receiver. This, in turn, requires a larger number of quantum wells in the modulator, which leads to a higher pre-bias and voltage swings. To mitigate this, either laser power can be increased or the transmission loss can be reduced. However, both these factors have limits. An effective solution would be to reduce capacitance at the transmitter and the receiver. For example, reducing this capacitance to 10fF reduces the optimum voltage swing to be within the stipulated supply voltage for the 65nm technology node.

Finally, we examined the impact of technology node scaling on the modulator design optimization. We observed that the modulator design metrics are relatively insensitive to the transistor performance and the power penalty (defined earlier as the difference in the total link power dissipation at the available supply voltage and that at the optimum supply voltage) has two opposite trends depending on the bit rate and the device capacitances. Lower bit rate and smaller device capacitances decrease the power penalty with technology scaling; while, the power penalty increases at higher

bit rates and larger capacitances. For the 32nm technology node with the ITRS recommended bit rate, (>30Gb/s), the power penalty becomes over 50%. For very high bit rates, (>30Gb/s) recommended beyond the 32nm technology node, electroabsorption modulated laser (EML) offers an attractive means to alleviate package complexity, reducing power dissipation. In addition, we also track the optimized laser power needed as a function of bit rate and device capacitances. For example, we find that the laser power becomes about 30% of the total link power for the 32nm technology node.

For on-chip applications, we compared CNT and optical interconnects with Cu interconnects using both commonly used metrics: latency and power and a novel compound metric, which captures system requirements more efficiently. This metric is defined as the bandwidth density per latency per power. This metric is motivated by the fact that larger bandwidth and a smaller latency can be obtained using more area resources. Hence area normalization is necessary in the form of bandwidth density. In addition, the total power budget can also be used to increase the aggregate bandwidth, making power normalization also imperative. In the future, because of multi-core architectures, the designers care about the bandwidth density, latency and power dissipation of global communications. We extensively examined the impact of device parameters-modulator and detector capacitances for optics, materials parameters-mean free path and packing density for CNTs, and system parameters - global clock frequency and switching activity, on both commonly used and the compound metrics. We find that at the 22nm technology node small detector and modulator capacitances for optical interconnects (~10fF) yields superior, at least comparable, performance with CNTs (practical, electron mean free path of 0.9μm) and Cu for greater than 35% and 20% switching activity, respectively. However, improving the mean free path of CNTs (~2.8μm) increases this crossover switching activity to 80%.

7.2 Future Recommendation

We designed a prototype electrical and optical off-chip link including the necessary circuits on a single chip in the 90nm technology node. Unfortunately, the process variations shifted the internally generated DC bias, which controls the current driving capability of the electrical transmitter. This is a critical component for testing the chip. This voltage shift resulted in driver having an insufficient strength to drive the network cable; hence, we could not get the signals, indicating the pass and fail of the communication transceivers. However, we overcame the significant engineering challenges in building a single chip electrical and optical transceiver. These non-trivial obstacles include a special package for optical interconnect, chip on board (COB) packaging, which compromises the performance of electrical interconnect. As a future recommendation, a cavity tuning method with the specifically handled processing would be very useful for improving the performance of end-device (modulator/detector) for optical links. With our design techniques described in Chapter 5 and after fixing the sensitivity issue of internal DC bias generating circuit, a revised chip preserving the current designs can be taped-out and tested.

This work can continue in many different directions. New novel device and advanced circuit technologies enhance the performance while reducing power. The system's impact of these devices should be investigated. For off-chip electrical links, a comparison between uni-directional signaling (especially as it pertains to phase amplitude modulation (PAM-4)) and bi-directional signaling (useful in a pin-limited environment) would be useful. On the other hand, optical links have various novel options whose performances need to be evaluated in system's context. The innovations include non-TIA (transimpedance amplifier) based receiver designs

including integrating current and sampling receivers [1], as well as novel device structures, such as monolithically integrated modulator [2], metal-semiconductor-metal (MSM) photo-detectors [3][4], and optoelectronic switch [5] which has a built-in gain. These solutions will inevitably speed the insertion of optical interconnect for shorter distance applications. For on-chip interconnects, several factors including the repeater area, can limit the minimum allowed wire pitch, hence the maximum bandwidth density. Bandwidth density and power dissipation will become even more critical as the high-end microprocessor designs shift toward multi-core architectures in the future. It is possible to make further comparisons between Cu, CNT, and optics in the light of various synchronous and asynchronous architectures options such as wire and wave pipelining, currently being investigated.

7.3 References

[1] A. E. Neyestanak, S. Palermo, H. C. Lee, and M. Horowitz, “CMOS Transceiver with Baud Rate Clock Recovery for Optical Interconnects,” *Symposium on VLSI Circuits*, pp. 410-413, 2004.

[2] Y. H. Kuo, Y. K. Lee, Y. Ge, S. Ren, J. E. Roth, T. I. Kamins, D. A. B. Miller, and J. S. Harris, “Quantum-Confined Stark Effect in Ge/SiGe Quantum Wells on Si for Optical Modulator,” *Journal of Selected Topics in Quantum Electronics*, vol. 12, no. 6, pp. 1503-1513, 2006.

[3] D. Buca et. al., “Metal-Germanium-Metal ultrafast infrared detectors,” *Journal of Applied Physics*, vol. 92, no. 12, pp. 7599-7605, Dec. 2002.

[4] C. O. Chui, A. K. Okyay, and K. C. Saraswat, “Effective dark current suppression with asymmetric MSM photodetectors in Group IV semiconductors,” *Photonic Technology Letter*, vol. 15, no. 11, pp. 1585-1587, Nov. 2003.

[5] A. K. Okyay, A. J. Pethe, D. Kuzum, S. Latif, D. A. B. Miller, and K. C. Saraswat, “Novel Si-based CMOS Optoelectronic Switching Device Operating in the Near Infrared,” *Optical Fiber Communication Conference*, March, 2007.

Chapter 7: Summary and Future Recommendations

Appendix A

Optical Receiver Modeling

A.1 Introduction

The optical receiver modeling was done by Krishnamoorthy et al. [1] and further discussed in Pawan et al. [2]. Although more complicated receiver can probably provide the better performance, this particular configuration, the transimpedance receiver, was conductive to study future scaling trends because of its simplicity. In this section, we will summarize the power modeling of the same configuration discussed before. The transimpedance receiver consists of a photo-detector followed by a transimpedance front-end stage with an inverting amplifier with feedback, as schematically depicted in Fig.A.1. Subsequent gain stages are stacked after the front-end to amplify the signal to the supply voltage. A CMOS inverter is chosen to be the gain stage because it will provide the largest gain for a give drain current [1].

Appendix A: Optical Receiver Modeling

With the optical receiver described above, we have following three constraints. 1) A receiver has the sufficient bandwidth for the I/O data rate of various technology nodes, according to ITRS. 2) A receiver has a sufficient digital SNR (DSNR), directly dictating the signal transmission reliability through the bit error rate (BER). For our modeling purpose, we chose a DSNR of about 7.9, which corresponds to a BER of 10^{-15} for a random data sequence. 3) A receiver achieves an output voltage swing, equal to the supply voltage for the particular technology generation (V_{dd}). We designed three design variables, the width of the amplifier transistors, the feedback resistance of the front-end, and the number of gain stages, which satisfying these constraints. In order to design three variables with three constraints, we need following assumptions. 1) The static power is the major source of power dissipation in the receiver, directly dependent on the transistor width. 2) The PMOS of the inverter is assumed to be twice that of NMOS to compensate for the lesser intrinsic drive current. 3) The front-end with the feedback is self-biased at the half of supply voltage ($V_{dd}/2$). The subsequent gain stages are also assumed to be biased by the previous stages at the same voltage. 4) The transistor sizing of all the gain stages is assumed to be identical and equal to that of the front-end stage. It is possible to lower power by sizing subsequent gain stage smaller; however, the size is kept the same to ensure that process variations do not lead to large voltage offsets, which could cause a higher BER. 5) The pole at the input of the front-end is assumed to be the dominant pole, limiting the bandwidth of the receiver. This seems reasonable because of a large detector capacitance compared to the gate capacitance of the transistors. 6) The noise model used in this work was developed in reference [4]. The noise sources considered were the amplifier noise due to drain current, the thermal noise, and the dark current and leakage current noise. We ignore the shot noise, negligible compared to other noise sources. 7) The detector-induced dark current is assumed to be negligible and the responsivity of the photo-detector is taken to be 0.5A/W. 8) The feedback resistor, R_f, is accomplished using PMOS device. The use of an active PMOS has the advantage of low area and

capacitance as well as the ability to very the resistance over a wide range. With using assumptions and constraints described above, we can derive the equations determining three design variables, the width of the amplifier transistors, the feedback resistance of the front-end, and the number of gain stages.

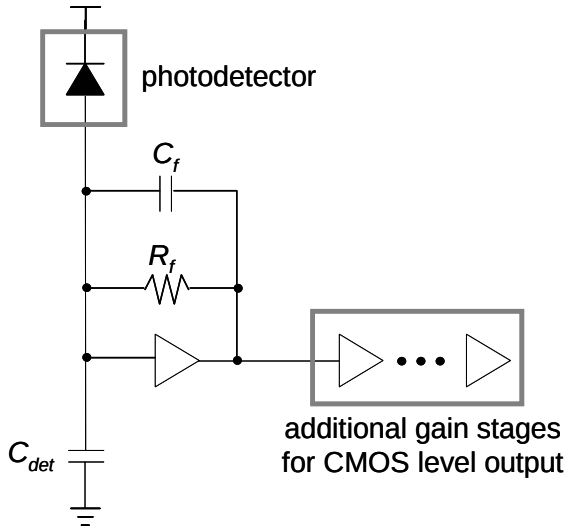


Fig.A.1 Schematic of the front-end with additional gain stages.

A.2 Bandwidth Constraint

Since we assumed the dominant pole at the input of the front-end, the bandwidth is simply inverse of the RC product where R_f is reduced and C_f is increased by a factor of A+1 because of Miller effect from the feedback impedance and it is given by [1]

Appendix A: Optical Receiver Modeling

$$f_{3dB} = \frac{A+1}{R_f C_T} = \frac{A+1}{R_f[C_{dec} + C_{inpg} + C_f(A+1)]} \tag{A.1}$$

Here, f_{3dB} is the 3dB bandwidth, R_f is the feedback resistance, A is the gain of the inverter with the feedback, C_T is the total capacitance looking into the input of the front-end and is the parallel combination of C_{det}, the detector capacitance, C_{inpg}, the input gate capacitance of the inverter, and C_f, the feedback capacitance. The input gate capacitance is the parallel combination of the gate-source capacitance of the NMOS and the PMOS. The gate-source capacitance in the saturation operation regime of the transistor is approximately given by sum of about two-third of the gate oxide capacitance. Consequently, the input capacitance is given by

$$C_{inpg} = C_{gsnmos} + C_{gspmos} = 3C_{gsnmos} = 3(\frac{2}{3}C_{ox}WL_{gate} + C_{ox}Wl_{overlap}) \tag{A.2}$$

Here, C_{gsnmos} and C_{gspmos} are the gate to source capacitances of the NMOS and the PMOS of the inverters, C_{ox} is the gate capacitance per unit are of the MOS transistor, L_{gate} is the transistor gate length, W is the width of the NMOS, and $l_{overlap}$ is the overlap of the gate over the source and drain regions due to lateral diffusion under the gate. This is taken to be about 30% of the gate length on one side and includes the effect of fringe capacitance. The feedback capacitance, C_f, is the sum of the gate-drain capacitance of the FETs and the capacitance of the feedback resistor. The gate-drain capacitance for the PMOS and NMOS in saturation region is simply the overlap capacitance of the transistors. Whereas, the relationship between resistance and the capacitance of the feedback resistor can be easily derived using triode operation equation of a PMOS transistor biased at V_{dd}. The total feedback capacitance is the parallel combination of these two capacitances and is given by

$$C_f = \frac{n}{R_f} + 3C_{ox} l_{overlap} W, \quad \text{where } n = \left(\frac{1.5 L_{gate}^2}{\mu_{effpmos}(V_{dd} - V_{th})} \right) \tag{A.3}$$

Here, V_{dd} and V_{th} are the supply voltage and the threshold voltage of the transistor respectively, $\mu_{effpmos}$ is the effective mobility of the PMOS. The gain with feedback, A, is simply the product of transconductance of the inverter and the parallel combination of inverter output resistance and feedback resistance as seen at the output.

$$A = g_m (R_o // \frac{R_f A}{A+1}) \tag{A.4}$$

Solving (A.4) for A leads to

$$A = \frac{R_o (g_m R_f - 1)}{R_o + R_f} \tag{A.5}$$

Here, g_m is taken to be the sum of the transconductance of NMOS and PMOS, whereas, R_o is considered to be the parallel combination of NMOS and PMOS output resistance.

By substituting (A.2), (A.3), and (A.5) into (A.1), we can explicitly solve for R_f in terms of the width of the transistor, W, at a given f_{3dB}, the detector capacitance, and the transistor parameters.

$$R_f = \frac{(1 - f_{3dB} \cdot n)(g_m R_o + 1) - f_{3dB} \cdot R_o (C_{\det} + C_{inpg})}{f_{3dB}[C_{\det} + C_{inpg} + 3C_{ox} W l_{overlap} (g_m R_o + 1)]} \tag{A.6}$$

Note here that C_{inpg} given by (A.2) and g_m, and R_o are also the function of W.

A.3 DSNR Constraint

The noise behavior of this transimpedance configuration has been studied in [3]. The total mean-square equivalent input noise due to the input devices of the receiver front-end arises from the Johnson noise of the feedback resistor, the photo-detector dark current, and the input FET gate leakage current, the 1/*f* or "flicker" noise in the FET's, and the channel noise of the input FET.

$$\left\langle i_n^2 \right\rangle \cong \frac{4kT}{R_f} I_2 B + 2qI_l I_2 B + 4kT\Gamma \frac{(2\pi C_T)^2}{g_m} f_c I_f B^2 + 4kT\Gamma \frac{(2\pi C_T)^2}{g_m} I_3 B^3 \tag{A.7}$$

Here B is the bit rate, I_l is the sum of the photo-detector and FET leakage current, I_2, I_3, and I_f are the normalized Personick noise bandwidth integrals, f_c is the 1/*f* noise corner frequency, and Γ is the excess channel noise factor term associated with short-channel transistors. However, the calculation of the Personnick integrals are strictly valid only for a family of rectangular input pulses and a raised cosine output pulses in a linear, equalized channel that is filtered prior to the decision circuit. The noise performance of an unequalized transimpedance receiver has been analyzed in [4] without making assumptions on the input and output pulse shapes. The mean square input noise current can be written as

$$\left\langle I_n^2 \right\rangle = \sigma_n^2 = \frac{4kT}{R_f} + 2qI_1\phi_1 + 4kT\Gamma\frac{(2\pi C_T)^2}{g_m}\phi_2 \tag{A.8}$$

Where,

$$\phi_1 = \frac{1+g_m R_o}{4[R_o(C_{in}+C_{out})+R_f(C_f+C_{in})+g_m R_o R_f C_f]} \tag{A.9}$$

$$\phi_2 = \frac{(1+g_m R_o)^2}{16\pi^2[R_o(C_{in}+C_{out})+R_f(C_f+C_{in})+g_m R_o R_f C_f][R_o R_f\{C_f(C_{in}+C_{out})+C_{\text{int}er}C_{out}\}]} \tag{A.10}$$

Here, k is the Boltzman constant, T is the temperature, q is the electronic charge, C_{in} is the sum of C_d and C_{inpg}, and C_{out} is the output capacitance of the front-end gain stage. This is given by the diffusion capacitance at the drain for both PMOS and NMOS in parallel with the load capacitance of the next stage. The diffusion capacitance is taken to be approximately half of the gate capacitance value [5]. Further, since the transistor sizing is assumed to be identical for all stages, load capacitance is the same as C_{inpg}.

Having the noise performance of the receiver front-end, we tackled the DSNR constraint in terms of current given by

$$DSNR = \frac{I_{on} - I_{off}}{\sigma_{on} + \sigma_{off}} \tag{A.11}$$

where, I_{on} and I_{off} are the average signal current for data "1" and "0", respectively, σ_{on} and σ_{off} are the root mean square current noise at the input for data "1" and "0", respectively. Because we assumed that the dark current is negligible compared to the

on current, the average current in the case of data "0" is zero. Also, since the shot noise is assumed to be negligible, the noise sources are approximately signal-independent. Hence σ_{on} is assumed to be approximately same as σ_{off}, thus (A.11) reduces to

$$DSNR = \frac{I_{on}}{2\sigma_{on}} \tag{A.12}$$

Here, I_{on} is given by

$$I_{on} = R_{sp} P_{bit} \tag{A.13}$$

Where, R_{sp} is the responsivity of the photo-detector and P_{bit} is the average optical power per bit.

(A.12) along with (A.8) and (A.13) gives us the second equation with two unknowns, R_f and W. We can solve for width by substituting (A.12) into (A.6). Subsequently, we can use (A.6) to solve for R_f. After we have followed this procedure, we would have obtained two design variables, the front-end transistor width and R_f.

A.4 Voltage Swing Constraint

A single transimpedance gain stage typically will not be sufficient to produce the voltage swings constraint. Also, to reduce the required swing at the input (and hence improve the sensitivity), additional stages of inverter amplifiers or gain-broadened inverters can be used. This allows smaller input photocurrents without

reducing the bandwidth of the receiver. The gain of these additional stages was simply the product of the output resistance and the transconductance of the inverter transistors.

$$A_A = g_m R_o \tag{A.14}$$

Now, we constrain the output of the receiver to be equal to V_{dd}. Thus, we have

$$V_{out} = V_{dd} = V_{front}(A)(A_A)^N \tag{A.15}$$

Here, N is the number of additional gain stages and V_{front} is the voltage swing at the front-end input. The product of V_{front} and A is the voltage swing at the output of the first transimpedance stage, where A being the gain of the first stage. The voltage swing at the front-end input (V_{front}) is dictated by the input resistance, R_f reduced by a factor of A+1 because of Miller effect from the feedback impedance, and input current (I_{on}) given by (A.13). Also as the operational frequency is comparable to the receiver bandwidth, this input resistance can be converted to f_{3dB}, 3-dB bandwidth taken to be 0.7 times bit rate (B). Hence,

$$V_{front} = \frac{I_{on}R_f}{A+1} = \frac{I_{on}}{2\pi f_{3dB}C_T} = \frac{I_{on}}{(2\pi C_T)(0.7B)} = \frac{I_{on}}{4.4C_T B} \tag{A.16}$$

From (A.15) and (A.16), we can obtain, N, the number of additional stages required as

$$N = \frac{\log(\frac{V_{dd}\,4.4C_T B}{AI_{on}})}{\log(A_A)} \tag{A.17}$$

If N is not an integer, then it is taken to be the next higher integer. With the number of gain stage from (A.17), and R_f and W from (A.6) and (A.12), we have all the variables for our design, which meets the DSNR, bandwidth, and the V_{dd} output swing constraints. Note the width obtained here only gives the minimum allowed with these constraints. However, the width can be larger than this minimum and the receiver can still meet the constraints. This is because as the width gets larger for the same R_f, the drain current noise reduces and DSNR improves and this fact is valid up to a point where the input gate capacitance of the transistor is the same as the sum of all other capacitances at the input [3][6]. However, the static power consumption tends to be very high at these large widths, thus, renders it unsuitable for short-distance, inter-chip and inter-board, application. We will use the fact that a larger width than the minimum calculated through above equations can still meet our constraints and further explore the possibility of total receiver power minimization.

A.5 g_m and R_o Calculation using Short Channel Equations

The transconductance (g_m) and the output resistance (R_o) of a transistor are needed for calculating R_f and W. These values can be easily calculated from the drain current equation by taking the partial derivative with respect to gate-source voltage and source-drain voltage. We start with the short channel equation for the drain current of an NMOS in the saturation regime.

$$I_d = \frac{\mu_{eff} C_{ox} W E_{sat} (V_{gs} - V_{th})^2}{2(V_{gs} - V_{th} + E_{sat}(L_{gate} - \Delta L))} \text{ for } V_{ds} > V_{dsat} \qquad \text{(A.18)}$$

Here, E_{sat} is the velocity saturation field. V_{gs}, V_{ds}, and V_{dsat} are the gate-source, drain-source, and saturation voltages respectively. ΔL is the reduction in the channel length because of pinch off after saturation and is semi-empirically given by [7]

$$\Delta L = l_{fit} \ln\left(\frac{V_{ds} - V_{dsat}}{E_{sat} l_{fit}} + \frac{E_m}{E_{sat}}\right) \tag{A.19}$$

Where,

$$V_{dsat} = \frac{(V_{gs} - V_{th}) l_{gate} E_{sat}}{V_{gs} - V_{th} + L_{gate} E_{sat}} \tag{A.20}$$

$$E_m = \sqrt{\left(\frac{V_{ds} - V_{dsat}}{l_{fit}}\right)^2 + E_{sat}^2} \tag{A.21}$$

and the fitting parameter l_{fit} is given by

$$l_{fit} = 0.22 x_{ox}^{\frac{1}{3}} x_j^{\frac{1}{2}} \tag{A.22}$$

Here, x_{ox} is the thickness of the gate oxide and x_j is the junction depth. Accounting for the deterioration in the mobility of the transistor due to vertical gate electric field, μ_{eff} is empirically given by [7].

$$\mu_{eff} = \frac{\mu_o}{1 + \theta(V_{gs} - V_{th})} \tag{A.23}$$

Here, μ_o is the surface mobility without vertical electric field and θ is a fitting

parameter. θ is calculated by using the drain current from ITRS along with (A.18)~(A.23). Both the gate-source and source-drain voltages were assumed to be V_{dd} for its generation at the ITRS on-current value.

From above equations, the transconductance of a transistor was calculated by taking the partial derivative of drain current with respect to gate-source voltage and we obtain

$$g_{mtrans} = \frac{\partial I_d}{\partial V_{gs}} = I_d\left(\frac{2}{V_{gs}-V_{th}} - \frac{1}{V_{gs}-V_{th}+E_{sat}(L_{gate}-\Delta L)} - \frac{\theta}{1+\theta(V_{gs}-V_{th})}\right) \tag{A.24}$$

Similarly, the output resistance of the transistor was calculated by taking the inverse of the partial derivative of the drain current with respect to source-drain voltage and we obtain

$$R_{otrans} = \frac{1}{\partial I_d / \partial V_{ds}} = \frac{E_m(V_{gs}-V_{th}+E_{sat}(L_{gate}-\Delta L))}{E_{sat}I_d} \tag{A.25}$$

Using (A.24) and (A.25), the g_m and R_o of the inverter were evaluated by adding the g_{mtrans} of both PMOS and NMOS and by obtaining the parallel combination of R_{otrans} of PMOS and NMOS, respectively. Note, both g_m and R_o have to be evaluated at the bias point, which is chosen to be at $V_{dd}/2$ in our design.

A.6 Power Calculation

From the previous sections, we have all the equations to calculate the design variables: the front-end transistor width, the feedback resistance, and the number of gain stages. The power per stage is the static power dissipation in a transistor biased at $V_{dd}/2$. Thus,

$$P_{perstage} = V_{dd} I_d \tag{A.26}$$

Where, I_d is given by (A.18) at $V_{dd}/2$ and is a linear function of the width, W. The total power of the optical receiver is the product of the number of all gain stages and the power dissipation per stage and we obtain

$$P_{receiver} = (N+1) P_{perstage} \tag{A.27}$$

Where, N+1 is the number of gain stages in addition to the first feedback stage, and is given by (A.17).

As we discussed in section A.3, the receiver power obtained using above equations use the minimum front-end transistor width. However, using this minimum width only guarantees the minimum power per stage, which may not translate to minimization of the total receiver power. For $R_o >> R_f$(width of the transistor is small), as the width of the transistor increases, although the power per stage goes up, there is a possibility that the gain of the front-end may also go up especially; hence, it may result in a reduction in the number of gain stages, thus, a possible decrease in total receiver power. However, this tendency has an impact on the reduction in the input voltage swing due to increase of input capacitance. This is especially true for small detector capacitance, hence the transistor gate capacitance contribute significantly to the total input capacitance, which may be the case of scaled technology. To account for the possibility of a lower total receiver power at a width larger than the minimum

width, we increase the width in small increments and calculate corresponding receiver power dissipation. Finally, we choose the width corresponding to the least power, and it ensures the absolute power minimization meeting the DSNR, bandwidth, and voltage swing constraints.

A.7 References

[1] A. V. Krishnamoorthy and D. A. B. Miller, "Scaling Optoelectronic-VLSI Circuits into the 21st Century: A Technology Roadmap," *Journal of Selected Topics in Quantum Electronics*, vol. 2, no. 1, pp. 55-76, Apr. 1996.

[2] P. Kapur, *Ph.D. Thesis,* Stanford University, 2002.

[3] R. G. Smith and S. D. Personick, "Receiver Design for Optical Fiber Communication Systems," in *Topics in Applied Physics*, H. Kreseel, Ed. New York: Springer Verlag, vol. 39, 1982.

[4] J. J. Morikuni, A. Darchoudhury, Y. Leblebici, and S. M. Kang, "Improvement to the Standard Theory for Photoreceiver Noise," *Journal of Lightwave Technology*, vol. 12, no. 4, pp. 1174-1184, July 1994.

[5] R. Ho, K. W. Mai, and M. A. Horowitz, "The Future of Wires," *Proceedings of the IEEE*, vol. 89, no. 4, pp. 490-504, Apr. 2001.

[6] B. I. Kasper, "Receiver Design," in *Optical Fiber Communication*, S. E. Miller and I. P. Kaminow Eds. New York: Springer Verlag, pp. 689-722, 1982.

[7] S. S. Wong, *EE316 class notes*, Stanford University, 1997.