Optimizing Nanophotonics: from Photoreceivers to Waveguides

Christopher Lalau Keraly

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2017-20
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-20.html

May 1, 2017

Optimizing Nanophotonics: from Photoreceivers to Waveguides

by

Christopher Lalau-Keraly

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eli Yablonovitch, Chair
Professor Ming Wu
Professor Feng Wang

Summer 2016

Optimizing Nanophotonics: from Photoreceivers to Waveguides

Abstract

Optimizing Nanophotonics: from Photoreceivers to Waveguides

by

Christopher Lalau-Keraly

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Eli Yablonovitch, Chair

Optical communication systems are replacing electrical interconnects on shorter and shorter scales, thanks to the large bandwidth they can provide and their better energy efficiency over long distances. Optical circuit boards or even on-chip interconnects are becoming an increasingly attractive possibility, thanks to tighter integration of photonics and electronics in technology platforms such as Silicon photonics. Nevertheless in order for optical links to become competitive with their electrical counterparts at these very short length scales, their energy efficiency must still be drastically improved. State of the art systems today consume $\sim$1pJ/bit of energy to communicate information, which is orders of magnitude above theoretical bounds.

In this thesis, the discrepancies between the theoretical limits and real world performance are explored, with a focus on the photoreceiver, which dictates the sensitivity and therefore much of the energy used by the link.

A thorough modeling of optical links is performed, leading to the determination of optimal receiver circuit topologies to improve the sensitivity and reduce the power consumption of photoreceiver systems. This enables the identification of crucial performance bottlenecks and the establishment of a technological roadmap for future generations of optical interconnects.

Additionally an extremely efficient shape optimization technique using the adjoint method for passive nanophotonics is presented, in order to provide lower loss components thereby also offering a path to improve the performance of optical links.

This thesis would not have been possible without the help and support of many people, both from the scientific perspective and the personal one.

First of all I would like to thank Professor Eli Yablonovitch for being an inspiring scientist always pushing for a fundamental understanding of topics that at first sight may seem like they are more of an engineering challenge than a scientific question, as well as being a kind advisor who gave me more freedom in the way I went about my research than I could have hoped for.

I would also like to thank the other professors I had the privilege of working with: Professor Ming Wu, who encouraged me and provided me with invaluable guidance while I explored phototransistors, and Professors Vladimir Stojanovic and Elad Alon for their help with the more practical aspects and limitations of photoreceiver circuits.

Maybe one of the greatest lessons from this work is that Aristotle was spot on when he said "the whole is greater than the sum of its parts". I would never have been able to accomplish this work without the collaborations that emerged during my time at Berkeley, and I would like to thank the different graduate students with whom I had the privilege to work closely: Owen Miller, Samarth Bhargava and Andy Michaels on nanophotonics optimization, Ryan Going on phototransistors and teaching me the ways of the nanolab, Krishna Settaluri on demystifying photoreceiver circuits, and Kevin Messer for all those grueling quantum physics homeworks. I would also like to thank all the other Yablonovitch group members for countless hours of fascinating physics discussions.

I am very thankful to both Dr. Josephine Yuen and Shirley Salanio for being so kind despite my phobia of deadlines.

I would like to thank all the other people who made my time at Berkeley amazing: the wonderful roommates of the Palace (and it's annex of course), the Santa Barbara crew, the fantastic four and extensions, CSC, and the majestic Pacific Ocean.

Finally my greatest thanks go to my parents, whom I love more than anything.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The challenge of communications

With the rise of internet and of computing in general, the need for communications at all levels is rising exponentially from transoceanic data links to interconnects on chips.

According to [1], annual global IP traffic will pass the zettabyte (10^{21}) by the end of 2016. With ever more data intensive applications such as the advent of virtual reality and the democratization of mobile internet, the growth of bandwidth demand has no end in sight. The infrastructure required to support such traffic is growing at the same rate, and while the hardware itself is becoming more efficient and requiring less energy per bit communicated, the energy efficiency is not scaling as fast as the demand and therefore the global energy consumption of the internet is rising. The internet infrastructure in the USA is estimated to have reached 2% of total electrical use [2], and is expected to continue its energy consumption growth.

At the chip level itself, data communication is increasingly becoming a bottleneck for computing. While the computing speed and density of chips has vastly improved with the scaling of transistors, the ability to bring data onto the chip has not been able to follow the trend, and the gap between memory bandwidth and computation speed is growing, such that increasing clock speed does not provide the same benefits as it used to. Additionally, one could argue that the ever increasing power consumption on chips also stems from communications: indeed modern microprocessors have up to 14 layers of metals for interconnects that need to be charged and discharged to move data around.

As one can see, communications is presenting an increasingly great challenge at many different levels. If we wish to continue to improve our computation and communication ability the way we have in the past, new solutions must be found that can provide both high bandwidths and low energy consumption.

1.2 The conquering growth of optical links

From the first transatlantic optical communication link (TAT-8 in 1988) to the first experimental demonstration of data being communicated to and off chip by light [3], optical communication systems have been gradually replacing copper wires over shorter and shorter lengths. State of the art commercially available systems today are replacing copper wires up to a few meters long. The emergence of nanophotonics and most notably Silicon Photonics technology [4], where optical components such as waveguides, modulators and detectors are manufactured in the same process and on the same chips as transistors brings the promise of huge improvements to current optical communications systems.

Two major reasons explain the replacement of copper wires at longer lengths: higher bandwidth and energy efficiency. Indeed the carrier frequency of light is $\sim$ 200 Thz, theoretically allowing huge amounts of data to be sent over optical fibers. The practical implementation of this corresponds to the use of dense wavelength division multiplexing (DWDM), where several channels are communicated with different wavelengths serving as the orthogonal carriers. Optical fibers today can have losses less than a dB/km, compared to RF cables which have losses of 10's of dB/km, meaning that they can propagate over much larger distances, or alternatively need to launch a lot less energy per bit for the same received power.

1.3 Future objectives for optical communications

At shorter scales, electrical links have yet to be replaced by optical ones. Indeed as the link distance becomes shorter, electric wires consume sufficiently low energy that it is not yet favorable to replace them with optics, where state of the art systems today consume roughly 1pJ/bit. Nonetheless there is still vast progress headroom for optical links, and the bandwidth they offer can still be massively larger than that of electrical connections. Indeed modern photoreceivers need $\sim$ 20 000 photons to accurately resolve every bit whereas physics dictates a quantum limit of $\sim$ 20, indicating a possible improvement of three orders of magnitude in the power burned on the transmitter side alone. Commonly cited objectives for chip to chip links range in the $\sim$ 100fJ per bit, and drop to $\sim$ 10fJ per bit when considering on-chip interconnects [5]. These energy requirements, when combined with the extremely high bandwidths needed pose a number of challenges for optical links. Ultimately, one could even hope that optical links might reach eve higher efficiencies and approach the Landauer limit ($kT \sim 4\times10^{-}21$ J/bit). The purpose of this work is to explore the reasons for such a disparity between current performance and theoretical limits, as well as offer solutions to approach these objectives and limits, with a focus on the performance of the photoreceiver.

1.4 Outline

Chapter 2 presents the basics of photoreceiver systems, chapter 3 calculates and presents the noise performance of different photoreceiving front ends. In chapter 4, a full model of an optical link is presented, and the co-optimization of the different parts is performed. Chapter 5 explores the concept of using phototransistors for optical communications and shows some of limitations of these types of devices, and chapter 6 presents extremely efficient shape optimization methods for passive components in electromagnetics.

Chapter 2

Photoreceiver basics

In this chapter, the basics concepts of direct detection photoreceiver systems are explained, without which it is impossible to quantify the performance of any given system. After a brief overview of the generic architecture of photoreceiver systems, the effect of bandwidth and noise on the photoreceiver performance will be covered. Finally photoreceiver devices will be presented.

2.1 Generic photoreceiver system

Coherent detection versus direct detection

Photoreceiver systems come in many different forms and flavors, depending on the specific application they are designed for. For example, coherent detection systems, which measures both the amplitude and phase information contained in the electric field of the incoming light, allow for a lot more information to be coded into the different degrees of freedom available. In order to be able to measure the phase of the incoming signal, coherent detection systems need an absolute phase reference which must be provided by a local oscillator phase locked to the incoming signal. This additional complexity comes at the cost of extra power burned in the receiver and means that coherent detection is mostly used where the receiver power consumption is not crucial to the system performance, such as satellite communication or long haul fiber optics systems like transoceanic systems.

Direct detection, on the other hand, relies on the measurement of signal energy. This greatly simplifies the detection scheme compared to coherent detection, as there is no longer any need for a local oscillator. While it is no longer possible to encode information on the entire quadrature plane, it is still possible to encode more than one bit per symbol, using schemes such as pulse amplitude modulation (PAM), albeit at a higher energy cost than for coherent schemes.

Since the goal of this thesis is to explore solutions for energy efficient short haul links, only direct detection with on-off-keying (OOK) signaling schemes will be studied.

Generic photoreceiver for OOK direct detection

The most general receiver system is composed of several components in order to successfully convert the optical input signal to a digital rail to rail signal. As illustrated in figure 2.1, the first component is the photodetector device, which converts the optical signal to an electrical one. This second stage performs amplification of the signal in the electrical domain, and is followed by an equalization stage. Finally a decision circuit takes the amplified and equalized electrical signal and decides whether the received bit is a 0 or a 1.

Naturally this is a schematic representation, and actual implementation may differ in various ways. For example, it is not impossible to forgo the amplification and equalization entirely, if the input signal is strong enough and properly shaped. Nevertheless it provides a general framework to study different photoreceiver architectures.

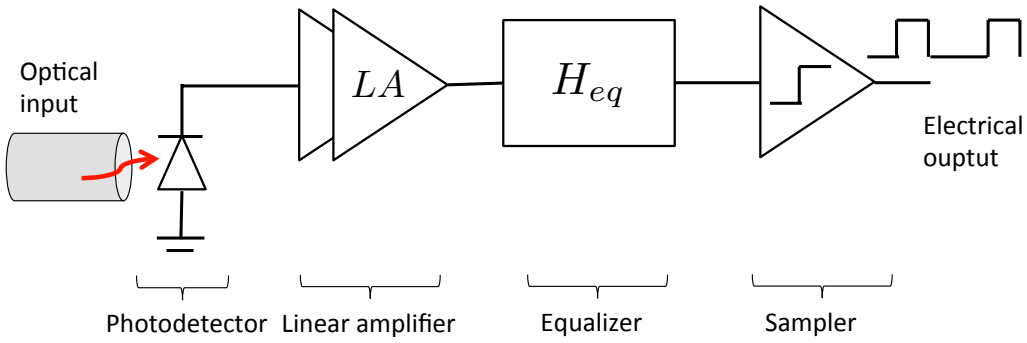

Figure 2.1: Photoreceiver system schematic

Figures of merit of a photoreceiver system

The ultimate goal of any communication link is to transmit information accurately and at speed, while using as little energy as possible. For a digital link, the speed is quantified by the datarate, while the accuracy is quantified by the bit error ratio (BER) i.e, the number of bit errors divided by the total amount of transferred bits. In most cases, the BER rate can be reduced by increasing the amount of signal power in order to overcome noise. This leads to another figure of merit which is the sensitivity, defined as the minimum input power required to achieve 10^{-9} BER. Sensitivity can be given in dBm, or equivalently in photons/bit. Finally the last important figure of merit is the total energy per bit required to transmit data, including the photon energy and the receiver energy.

2.2 Effect of bandwidth

The design of a photoreceiver must take a large number of constraints into consideration, one of the most important of them being it's bandwidth. For random binary data arriving at the receiver, the spectrum is [6]

$$S(f) = T_n\left[\frac{sin(\pi f T_b)}{\pi f T_b}\right]^2 \tag{2.1}$$

As we can see here, the bandwidth of the signal is large and goes all the way to DC. This puts a broadband requirement on the receiver itself, and we explore here the adverse effects of limiting this bandwidth.

Low and High pass filtering

Low pass

If the receiving system has a bandwidth that is too low compared to the data-rate, the low frequencies of the signal will be amplified more than the high frequency components. This affects the received data most when a run of similar bits happen (i.e when many ONEs or ZEROs follow each other in a row). Intuitively it makes sense that the run has a large "instantaneous" DC component, compared to an alternating stream of ONEs and ZEROs. This is illustrated in figure 2.2. As shown, at the end of a run, the DC signal value has drifted enough that it affects the following bit, causing inter-symbol interference (ISI).

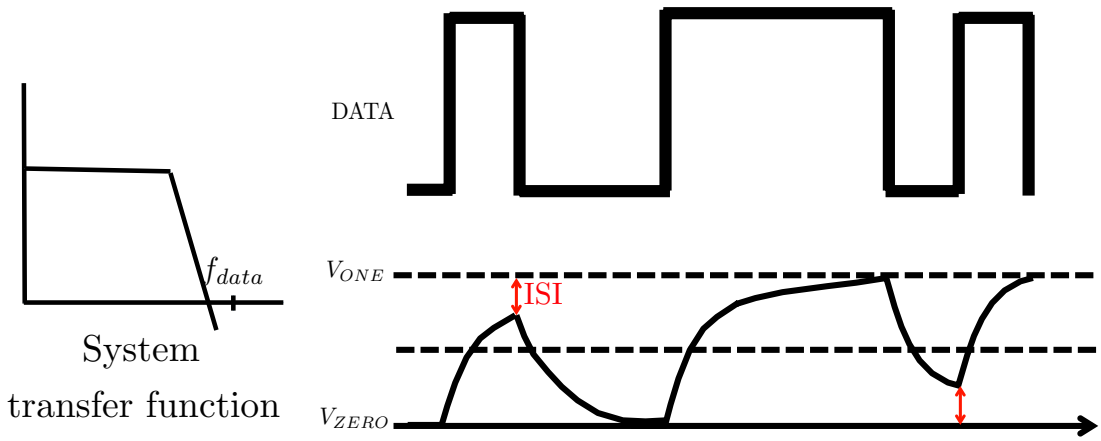

Figure 2.2: Effect of low pass filtering with Inter symbol interference (ISI) emphasized

high pass

The effect of high pass filtering is opposite: it attenuates the DC component of the signal, as illustrated in figure 2.3. High pass filtering is most common when AC coupling is used along this signal path, which will not be considered in this work.

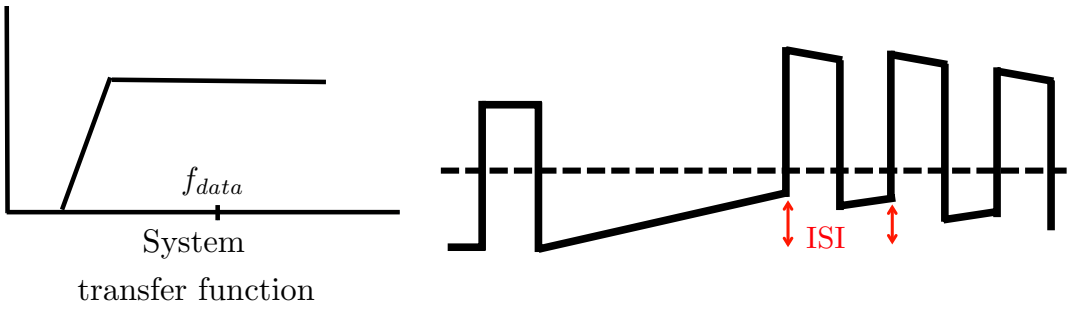

Figure 2.3: Effect of high pass filtering with ISI ephasized

Increasing the bandwidth of the receiver also increases the noise, as will be seen in the next section, so that making the bandwidth as large as possible is not necessarily the optimal solution.

2.3 Effect of noise

Noise is one of the most critical problems that hinders the performance of photoreceivers, and is ultimately one of the primary limiting factors when it comes to improving the sensitivity of a photoreceiver. A good understanding noise sources and how to model them appropriately is crucial if one wants to be able to quantify the sensitivity of a given photoreceiver.

Fundamentals of noise modeling

Noise can be defined as any signal present at the receiver other than the desired one. Noise sources are numerous, and can come from a wide variety of effects, from laser noise, to mode mixing, to electronic noise. Focus here is set on the electronic noise coming from the receiver.

If the noiseless signal received at the decision circuit is S(t), and all the noise sources along the way add an extra component n(t), the decision circuit actually sees a signal S(t)+n(t), as illustrated in figure 2.4

When the decision circuit decides if the bit received is a ONE or a ZERO, it compares the signal voltage to a reference voltage and outputs a ONE if it is above and a ZERO if

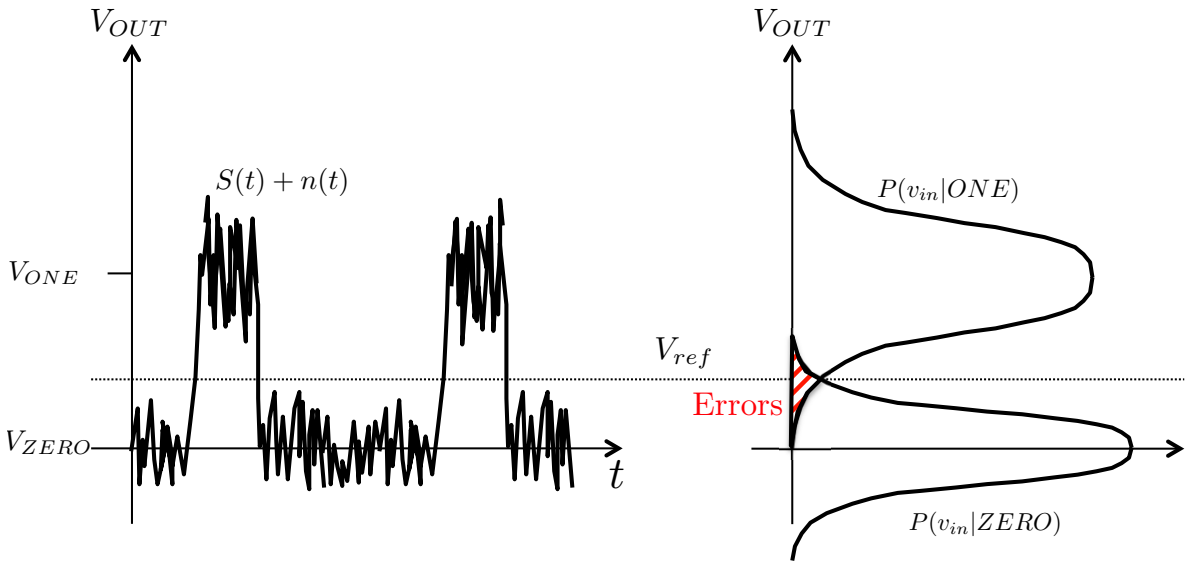

Figure 2.4: Effect of noise on signal, and illustration of the noise probablity distribution. V_{ref} is the reference voltage used by the decision circuit to decide whether a ZERO or a ONE is received. In this case, the ONEs are considerably more noisy than the ZEROs, which would indicate large photon shot noise

it is below. If the amplitude of the noise is large, it can distort the signal strongly enough that a wrong decision is made, and a bit error occurs. The BER is therefore a measure of the probability of the noise overwhelming the signal and leading to a wrong decision. The inherent randomness of noise means statistical analysis must be used and the probability density function (PDF) of the noise distribution is needed to calculate the occurrence of wrong decisions.

Noise PDF and SNR

As will be covered in more detail, most noise sources can be modeled as Gaussian noise and therefore have normal PDF $F(n)$ which can be written in terms of their variance σ as:

$$F(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{n^2}{2\sigma^2}} \tag{2.2}$$

$$\sigma^2 = \frac{1}{T}\int_0^T n^2(t)dt \tag{2.3}$$

where T is an arbitrarily long time. If the signal is in the voltage domain, σ is equivalently called the total integrated noise voltage V_n, and is just the root-means-square (RMS) value of the noise voltage. The probability at any time that the $n(t)$ is within $\frac{\Delta n}{2}$ of n_0 is

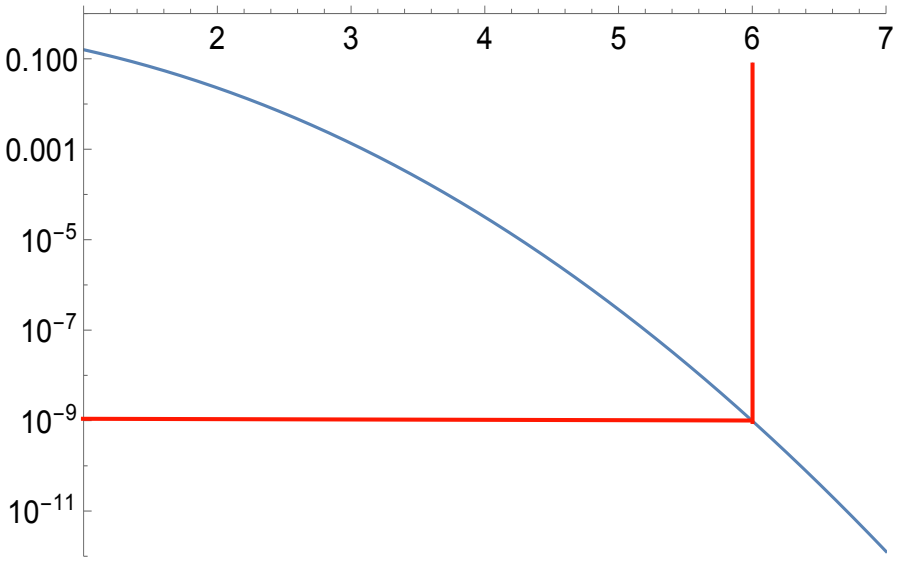

Figure 2.5: Plot of the Q function, which will give the current SNR required to achieve a certain gain

$F(n_0)\Delta n$.

When the noise is similar whether the bit is a ONE or a ZERO (which is usually the case if the noise is not dominated by photon shot noise) and the data is random, the optimal reference voltage V_{ref} is the mean of the signal voltage for a ONE and a ZERO, as illustrated in figure 2.4. The probability of the noise inducing an error on the system can then be calculated For a ZERO, if the signal voltage is 0V, an error occurs if the voltage induced by the noise is greater than V_{ref}, and that probability can be calculated by the PDF.

$$P_{error,ZERO} = \int_{V_{ref}}^{\infty} F(n)dn = Q(\frac{V_{ref}}{V_n}) \tag{2.4}$$

$$\text{where} \quad Q(z) = \frac{1}{2\pi}\int_{z}^{\infty} e^{-x^2/2}dx \tag{2.5}$$

The Q function defined here is plotted in figure 2.5.

Since in our case we have assumed an equal noise distribution for ONEs and ZEROs, we can easily conclude that the probability of error is the same for both and we finally have:

$$BER = Q(\frac{V_{ref}}{V_n}) = Q(\frac{V_{one}}{2V_n}) = Q(SNR) \tag{2.6}$$

where we have defined

$$SNR = \frac{V_{one}}{2\sigma} = \frac{V_{avg}}{V_n} \tag{2.7}$$

The SNR required for a BER of 10^-9 can be deduced from figure 2.5 to be 6.

Noise power spectral density (PSD)

We can now easily determinate the BER and sensitivity if we are able to determine the RMS value of the noise V_n. It is a lot easier to work in the frequency domain rather than the time domain when trying to determine the noise of a system. The noise PSD $v_n(f)$ is a function of frequency and represents the amount of noise power present at that frequency. It's units are therefore $V/\sqrt{Hz}$, and for a given noise PSD, we can calculated the total noise voltage as:

$$V_n^2 = \int_0^\infty v_n^2(f)df \tag{2.8}$$

It is especially convenient to use noise in the frequency domain, since we can easily calculate how the noise is shaped after any circuit, by multiplying $v_n(f)$ by the circuit's transfer function itself $H(f)$

Sources of noise and their modeling

Shot noise

Shot noise is an intrinsically quantum effect, and comes from the fact that current flowing is not continuous but is composed of discrete particles carrying $q = 1.6\ 10^{-19}C$ of charge. When electrons have to go over a potential barrier, the process has no memory: i.e it does not depend on how many electrons may have gone over it earlier. This results in Poissonian statistics for current flow. Shot noise is also present in the photon flux, for the same reason: photons are discrete, and their generation is random, so they follow the same statistics.

If an average number of events in a time period is λ, Poisson statistics dictate that the probability of n events actually happening is

$$P_\lambda(n) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{2.9}$$

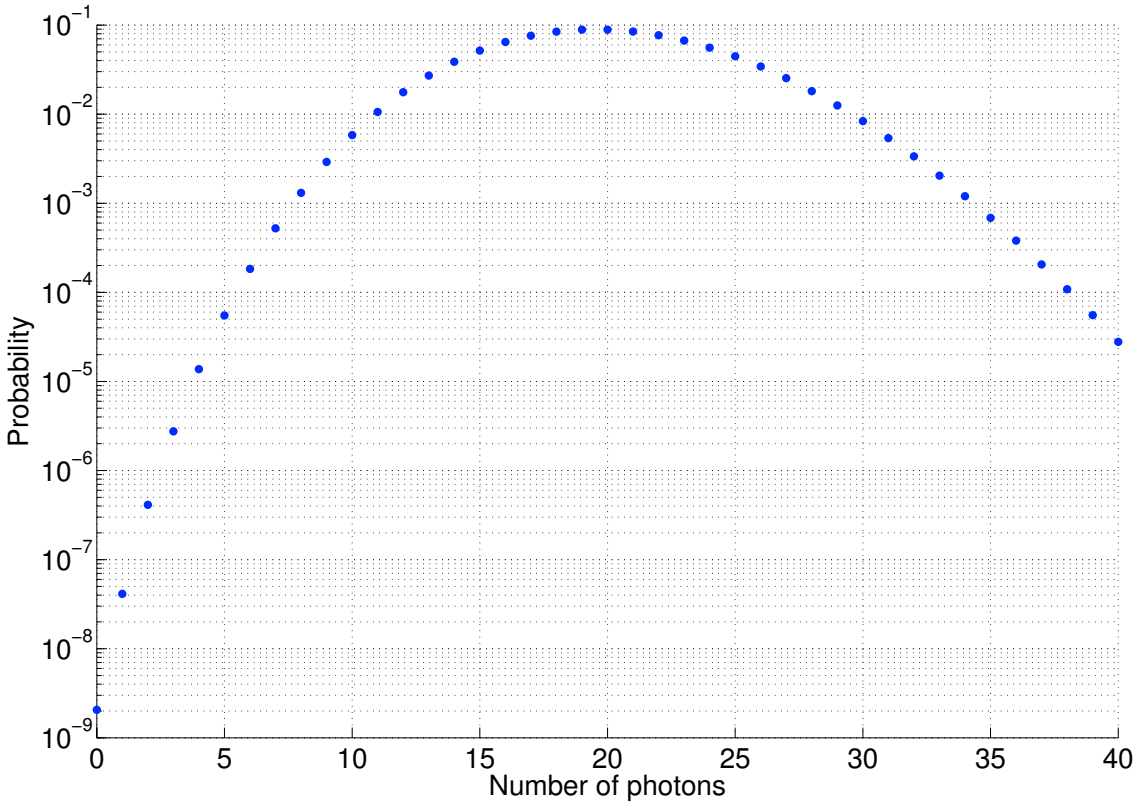

Figure 2.6: Probability of seeing a certain number of photons when the expectation value is 20, and the photons follow a Poisson distribution

This is plotted on figure 2.6, for $\lambda = 20$. We can see that the probability of not having any events happening is $\sim 2 \times 10^{-9}$, which is why 20 photons per bit for a ONE is usualy quoted as the "quantum limit". Indeed, if one is limited by the quantum shot noise, and able to accurately detect single photons, errors can only happen for false ZEROs, when no photon is received for a ONE. Given an equal proportion of ONEs and ZEROs, the probability of an error is therefore $\sim 1 \times 10^{-9}$ per bit

For values of λ large enough, Poisson distributions can accurately be approximated by a Gaussian distribution with the same variance, with a noise current PSD of:

$$i^2_{n,shot}(f) = 2qI \tag{2.10}$$

Shot noise is white, meaning it covers the entire spectrum, and is present for the base and collector current of bipolar transistors, for diode currents and for photon currents, among others.

Johnson noise

Johnson noise, or thermal noise, is caused by the random thermal fluctuation of electrons in a conductor. Just as shot noise, it is also white. It can be represented as a current source in parallel with the conductor with a noise current PSD of

$$i^2_{n,Johnson} = \frac{4k_b\Theta}{R} \tag{2.11}$$

Where Θ is the ambiant temperature in Kelvin. It is present not only on resistors but also in the channel of MOSFET transistors. In that case the PSD is

$$i^2_n = 4k_b\Theta\gamma g_m \tag{2.12}$$

where g_m is the transconductance of the MOSFET.

2.4 Photodetection devices

Optical absorption of semiconductors

The basis behind virtually all photoreceivers is based on the creation of electron and hole pairs in semiconductors when photons are absorbed in the material. Since electron and hole pairs are formed when electrons are excited from the valence band to the conduction band, to first order only photons with energies above the semiconductor bandgap energy can be absorbed, which explains why the onset of absorption is so abrupt in figure 2.7. The absorption tails (namely the Urbach tails) observed are due to disorder effects and phonons, which are not strong enough to provide large enough absorption coefficients for practical purposes.

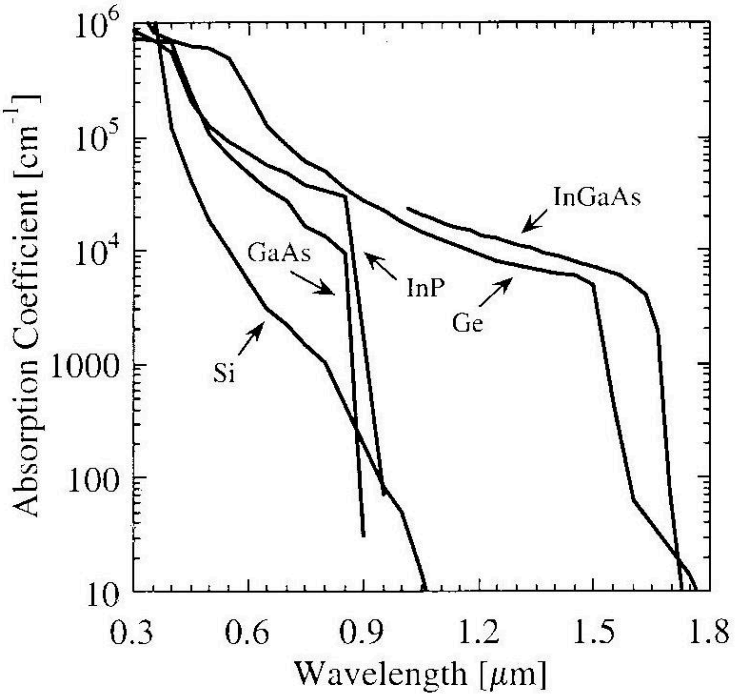

Figure 2.7: Optical absorption in common semicondutors, from [7]

From an energy perspective it is necessary for the photons to have a net energy above the bandgap, but from a momentum perspective, if the bandgap is indirect, such as in Silicon or Germanium, there can be no absorption without the extra help of a phonon. Nevertheless these processes are not rare, and Germanium can be used to absorb photons with energies below the direct bandgap, although it's absorption coefficient is rather small, with an absorption length of $\sim 20\mu m$.

p-i-n devices

The most commonly used semiconductor device to detect light is the p-i-n photodiode, which is a simple p-n junction with an undoped region between the p-doped and n-doped semiconductor. The band diagram of a reverse biased junction photodiode is illustrated in figure 2.8.

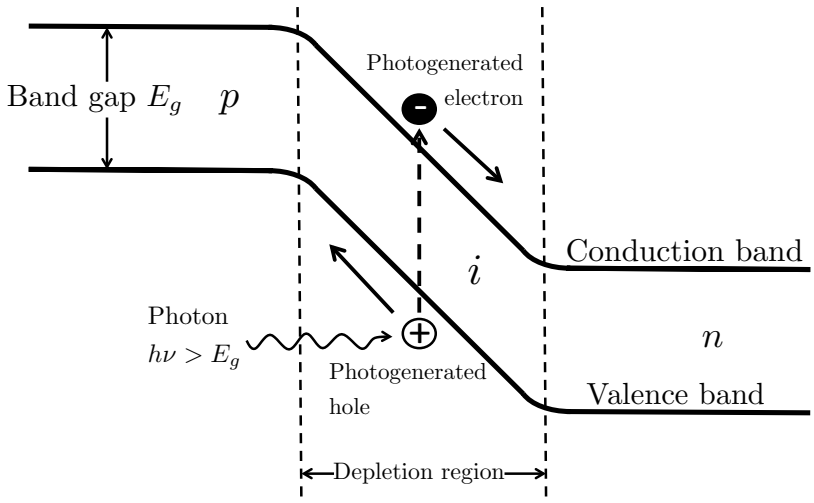

Figure 2.8: Energy band diagram of a reverse biased p-i-n junction

The electrons and holes generated by photons absorbed in the intrinsic region are quickly swept away by the electric field, and contribute one unit of charge to the current. The responsivity (ratio of current out to power in, in Amps/Watt) is therefore dictated solely by the quantum efficiency η of the device, according to

$$R = \eta \frac{q}{h\nu} \sim \eta \frac{\lambda}{1.24} \tag{2.13}$$

The intrinsic speed response of a photodiode comes from the time required for the generated photocarriers to transit through the intrinsic region of the device. In normal operation conditions, the bias is strong enough that the carriers will drift in velocity saturation regime. Since the slowest carriers are usually the holes, the transit time is

$$\tau_t = \frac{l}{v_h} \tag{2.14}$$

where l is the length of the intrinsic region. The exact bandwidth achieved by the device is subject to a few subtleties. If the illumination is uniform in the intrinsic region, the 3-dB bandwidth can be approximated as [8]:

$$f_{3dB} = \frac{0.44}{\tau_t} \tag{2.15}$$

In most practical cases the photodiode will be connected to a load resistor, and the bandwidth will not be limited by the transit time but by the RC time coming from the capacitance of the photodiode and the load resistor. The capacitance of a photodiode is

$$C = \frac{\epsilon_0 \epsilon_r A}{l} \tag{2.16}$$

where A is the transverse area of the diode.

The noise at the output of a photodiode is shot noise coming from both the photocurrent, and the dark current.

$$i^2_{p-i-n} = 2q(I_{ph} + I_{dark}) \tag{2.17}$$

This is usually not the dominant noise source though when a photodiode is used, as the load resistance will usually have much higher Johnson noise. This will be studied in the receiver architecture section.

Avalanche photodiodes (APD)

The basic structure of APDs is similar to that of p-i-n photodiodes, but the bias applied to the junction is large enough to cause avalanche of carriers: photo-generated electrons (or holes) gain a kinetic energy higher than that of the bandgap while drifting through the intrinsic region and are able to excite electrons from the valence band to the conduction band through impact ionization, thereby creating a new electron hole pair which in turn can cause more impact ionization. This means that there is an intrinsic gain mechanism in the device, as illustrated in figure 2.9. This is particularly attractive because the responsivity can be much larger than that of a simple photodiode. In most practical implementations of APDs, the absorption and multiplication regions are acturally separate, allowing for a better control of the different depletion widths.

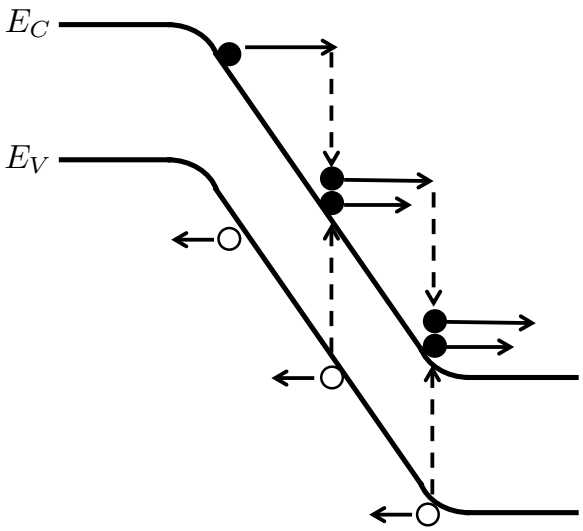

Figure 2.9: APD band diagram and multiplication process illustration

APD theory have been thoroughly documented elsewhere[8], and we will only cover the most important results here. The gain factor (or multiplication factor) is usually referred to as M, resulting in a responsivity for APDs of

$$R = \eta \frac{q}{h\nu} M \sim \eta \frac{\lambda}{1.24} M \tag{2.18}$$

While the average multiplication for every absorbed photon is M, is process itself is stochastic, and leads to excess noise at the output of the APD. An important factor in determining the excess noise is the ionization factor ratio between holes an electrons: $k = \alpha_p/\alpha_n$. Indeed if only electrons can create new electron hole pairs, the multiplication process is much better controlled and less variable than if both holes and electrons can cause ionization. In the case of photo-electron injection (meaning the multiplication region is on the n side of the absorption region) the excess noise factor can be calculated as [8]:

$$F_n(M) = kM + (1-k)(2 - \frac{1}{M}) \tag{2.19}$$

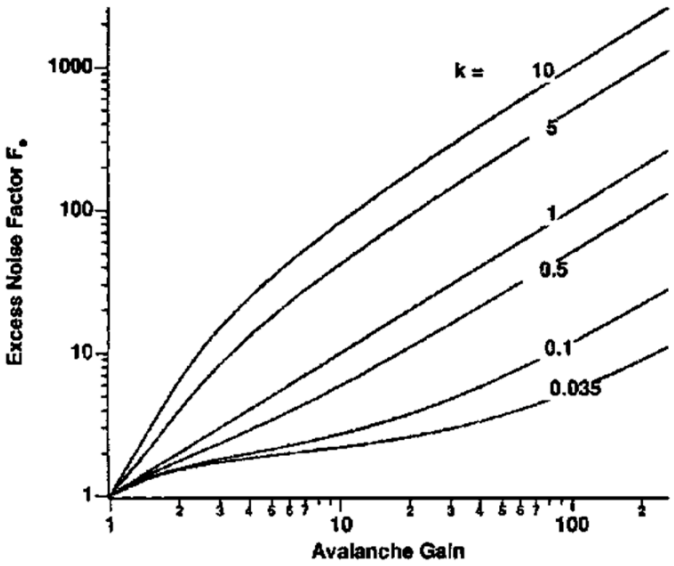

Figure 2.10: Excess noise factor in avalanche photodiodes, reproduced from [8]

The excess noise factor is just a multiplicative in the shot noise of the current running through the APD, so that the PSD can be written as.

$$i^2_{s,APD} = 2q(I_{ph} + I_{dark})F(M) \tag{2.20}$$

The k factor depends on the semiconductor in which the multiplication takes place. For example it can be as low as 0.02 in Silicon, which leads little excess noise. Most other material have larger ionization ratios, such as Germanium at 0.9, bringing about much higher noise. This has lead to the development of APDs [9] with an absorption region in Germanium and the multiplication region in Silicon, to benefit from the high absorption of Germanium and the excellent multiplication properties of Silicon.

Unfortunately the device itself requires a large bias voltage in order to cause impact ionization, adding additional complexity to the system, along with higher power consumption and reliability issues.

Phototransistors

Phototransistor are explored in chapter 5

Chapter 3

Noise calculations for different front ends

Noise calculations in modern circuit design are all performed via the use of CAD simulation software. While these tools are absolutely crucial for practical design, they do not necessarily give their user an deep understanding of the different trade-offs and limits of a given system. It is therefore crucial to be able to perform these calculations analytically to a reasonable level of accuracy. In this chapter, we present the general frame used by Personick [10] to calculate the noise performance photoreceiver systems, and we apply it to different front end topologies.

3.1 Input refered noise

In order to be able to perform a fair comparison of noise sources, it is important to refer them to the same node of the system. Indeed, two noises of equal magnitude will not have the same effect if one affects the signal before any gain is achieved, whereas the other comes after several gain stages. In the second case the signal has been amplified and is more immune to noise. Additionally, different noise sources will experience different spectral shaping depending on their location in the signal path. In order to be able to compare the different noise sources and quantify their effect with respect to the signal, it is necessary to refer them to the same node in the signal path. The most natural point is at the input, so that it can be compared directly to the photon current arriving at the photodiode.

$H(\omega)$ is defined as the transfer function of the front end of the amplifier being studied.

$$V_{out} = H(\omega) I_{in} \tag{3.1}$$

The noise, referred to the output of the front end is written as

$$V_N = H_N I_N \tag{3.2}$$

where I_N is a current noise source.

If subsequent amplification and equalization stages after the front end are assumed, with a transfer function $H_{eq}(\omega)$, such that the entire transfer function of the amplifier is $H(\omega)H_{eq}(\omega)$ we can write [1]

$$< V^2_{noise,out} >= \frac{1}{2} I_N^2 \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| H_N(\omega) H_{eq}(\omega) \right|^2 d\omega \tag{3.3}$$

$H_s(\omega)$ and $H_{out}(\omega)$ are defined as the input photon density and output voltage pulse transforms, respectively, so that

$$I_s(t) = \sum_{-\infty}^{+\infty} b_k \ \eta \ q \ n_{ph} \ h_s(t - kT) \tag{3.4}$$

$$V_{out}(t) = \sum_{-\infty}^{+\infty} b_k \ \eta \ n_{ph} \ h_{out}(t - kT) \tag{3.5}$$

where b_k is the value of the k^{th} bit (0 or 1), η is the absorption efficiency of the device, n_{ph} is the number of photons per bit arriving on the device, T is the bit duration, and with the following normalizations: $\int h_s(t)dt = 1, h_{out}(0) = 1$ (This of course implies an arbitrary normalization of the output pulse), and $h_{out}(kT) = 0$ for all k≠0, (which means we have chosen an output waveform that minimizes inter-symbol interference, such as a raised cosine).

We can therefore write

$$\frac{H_{out}(\omega)}{q \ H_s(\omega)} = H_{eq}(\omega) H(\omega) \tag{3.6}$$

And (3.3) is rewritten as

$$< V^2_{noise,out} >= I_N^2 \frac{1}{4\pi} \int_{-\infty}^{\infty} \left| H_N(\omega) H(\omega)^{-1} \frac{H_{out}(\omega)}{q \ H_s(\omega)} \right|^2 d\omega \tag{3.7}$$

And develop

$$\left| H_N(\omega) H^{-1}(\omega) \right|^2 = A + B\omega^2 \tag{3.8}$$

[1] In order to be consistent with Personick's original derivations where noises sources contributions are spectrally integrated from -∞, but with the modern way of writting the amplitude of noise sources, where it is assumed the integration starts at 0 Hz, we introduce an extra factor 1/2 compared to [10], but keep modern noise amplitudes, such that $i^2_{Johnson} = \frac{4k\Theta}{R}$ instead of $\frac{2k\Theta}{R}$.

So that

$$< V^2_{noise,out} >= I^2_N \frac{1}{4\pi} \int_{-\infty}^{\infty} (A + B\omega^2) \left| \frac{H_{out}(\omega)}{q\ H_s(\omega)} \right|^2 d\omega \tag{3.9}$$

we also perform the change of variables and definitions:

$$y = \frac{T\omega}{2\pi} \tag{3.10}$$

$$H'_s(\omega) = H_s(\frac{2\pi\omega}{T}) \tag{3.11}$$

$$H'_{out}(\omega) = \frac{1}{T} H_{out}(\frac{2\pi\omega}{T}) \tag{3.12}$$

This means that we can define the following integrals that depend only on the pulse shapes h_S and H_{out}, and not on the absolute bandwidth.

$$I_2 = \int_{-\infty}^{\infty} \left| \frac{H'_{out}(y)}{H'_s(y)} \right|^2 dy \tag{3.13}$$

$$I_3 = \int_{-\infty}^{\infty} \left| \frac{H'_{out}(y)}{H'_s(y)} \right|^2 y^2 dy \tag{3.14}$$

These are called the Personick integrals. Examples of the values they take will be given in section 3.2.

and get

$$< V^2_{noise,out} > = \frac{1}{2} I^2_N T \int_{-\infty}^{\infty} (A + B \left(\frac{2\pi y}{T} \right)^2) \left| \frac{H'_{out}(y)}{q\ H'_s(y)} \right|^2 dy \tag{3.15}$$

$$= \frac{I^2_N T}{2q^2} \left[A\ I_2 + B\ I_3 \left(\frac{2\pi}{T} \right)^2 \right] \tag{3.16}$$

Since $V_{out}(t) = \sum_{-\infty}^{+\infty} b_k\ \eta\ n_{ph}\ h_{out}(t - kT)$, and given the assumptions on H_{out}, the (normalized) output for a ONE is $V_{out} = \eta n_{ph}$ and $V_{out} = 0$ for a ZERO. Given equation 2.7, the minimum number of photons per bit for a ONE can be expressed as.

$$n_{ph} = \frac{2SNR}{q\ \eta} \sqrt{\frac{I^2_{N_R} T}{2} \left[A\ I_2 + B\ I_3 \left(\frac{2\pi}{T} \right)^2 \right]} \tag{3.17}$$

This method can be used for all types of front ends. It should be noted that the use of equation 2.7 implies that the noise sources PSD are symmetrical for a ONE and a ZERO, which is true as long as the dominant noise source is not shot noise from the signal (the excess noise of an APD can be considered as signal shot noise). In the case where the shot noise is dominant, the factor 2 in equation 3.17 disappears.

3.2 The Personick integrals

In order to calculate the I_2 and I_3 Personick integrals from equation 3.13 and 3.14, we must know what transfer function the entire systems undergoes before reaching the decision circuit. Naturally, smart equalization can be used to minimize the values of the integrals, but these may require complicated circuits which can consume large amounts of energy. It is interesting to calculate these integrals in the simpler case of N chained first order amplifiers having a transfer function

$$H_{system} = \frac{G}{(1 + j\frac{\omega}{\omega_0})^N} \tag{3.18}$$

The bandwidth of such a system can be approximated as [6]:

$$B = 2\pi\omega_0 \frac{0.9}{\sqrt{N+1}} \tag{3.19}$$

For an OOK signaling system, and as long as the effective bandwidth of H_{system} is above the Nyquist rate $f_{data}/2$, the Personick integral can be approximated as:

$$I_2 = \frac{T_{bit}\ \omega_0}{2\pi} \int_{-\infty}^{\infty} \frac{1}{(1+y^2)^N} dy = \frac{T_{bit}\ \omega_0}{2\pi} \frac{\sqrt{\pi}\ \Gamma(N-1/2)}{\Gamma(N)} \tag{3.20}$$

N	1	2	3	4	5
$\frac{\sqrt{\pi}\ \Gamma(N-1/2)}{\Gamma(N)}$	π	$\pi/2$	$3\pi/8$	$5\pi/16$	$35\pi/128$

$$I_3 = \left(\frac{T_{bit}\ \omega_0}{2\pi}\right)^3 \int_{-\infty}^{\infty} \frac{y^2}{(1+y^2)^N} dy = \left(\frac{T_{bit}\ \omega_0}{2\pi}\right)^3 \frac{\sqrt{\pi}\ \Gamma(N-3/2)}{2\Gamma(N)} \tag{3.21}$$

N	1	2	3	4	5
$\frac{\sqrt{\pi}\ \Gamma(N-3/2)}{2\Gamma(N)}$	∞	$\pi/2$	$\pi/8$	$\pi/16$	$5\pi/128$

Given the fact that for larger values of N, the effective bandwidth of 3.18 becomes a smaller and smaller fraction of $2\pi\omega_0$, as shown in 3.19, and the condition on the total system bandwidth of the system being above the Nyquist rate, the Personick integrals rarely go far from unity in real systems. Therefore we will use $I_2 \sim I_3 \sim 1$ for further calculations, unless noted otherwise.

3.3 Resistor loaded p-i-n front end

In the case of a simple resistively loaded front end with a follow on transistor stage and a low impedance output (such as a cascode), such as depicted in 3.1, the sources of noise are the resistor Johnson noise $I_{N,R}$, the transistor Johnson noise $I_{N,T}$, and the signal shot noise $I_{N,S}$. The transfer function is simply:

$$H(\omega) = g_m Z_{out} \frac{R_l}{1 + jR_l C\omega} \tag{3.22}$$

where

$$g_m = 2\pi f_t C_{ox} \qquad \text{the FET transconductance} \tag{3.23}$$

$$C = C_{PD} + C_{ox} \qquad \text{the input capacitance} \tag{3.24}$$

The bandwidth of the front end is

$$B_{RC} = \frac{1}{2\pi R_l C} \tag{3.25}$$

and

$$V_{N,out} = (I_{N,R} + I_{N,S})H(\omega) + I_{N,T} Z_{out} \tag{3.26}$$

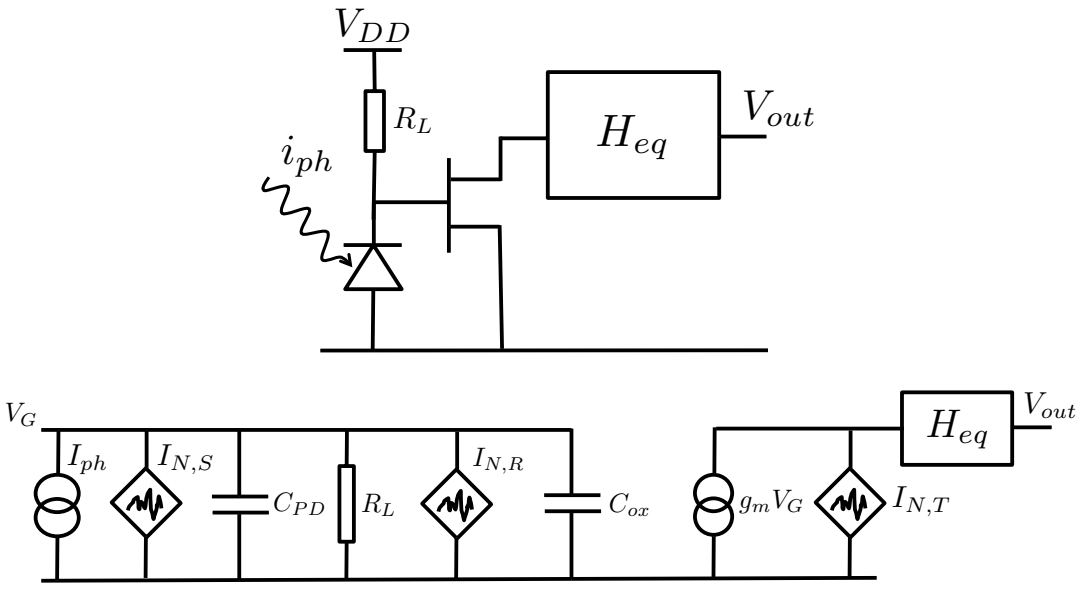

Figure 3.1: Resistor loaded photodiode schematic and small signal equivalent circuit

Using the method described in 3.2, we have:

$$n_{ph} = \frac{2SNR}{q\ \eta}\sqrt{(I_{N_S}^2 + I_{N,R}^2)TI_2/2 + I_{N,T}^2 T\frac{I_2 + I_3(R_l C 2\pi/T)^2}{2(g_m R_l)^2}} \tag{3.27}$$

where

$$I_{N,S}^2 = 2qI_{ph} \tag{3.28}$$
$$I_{N,R}^2 = 4k_b\theta/R \tag{3.29}$$
$$I_{N,T}^2 = 4k_b\theta\gamma g_m \tag{3.30}$$

which can be approximated as

$$n_{ph} \sim \frac{2SNR}{\eta}\sqrt{n_{ph}I_2 + 4\pi\frac{C}{6.4aF}\frac{B_{RC}}{f_{data}}I_2 + 4\pi\frac{C}{6.4aF}\frac{C}{C_{ox}}\frac{f_{data}}{f_T}\gamma I_3} \tag{3.31}$$

where the first term under the root is the shot noise contribution, the second term the resistor Johnson noise and the third term the transistor Johnson noise.

Regarding the signal shot noise, it is not symmetric for ONEs and ZEROs, which was one of they hypothesis in the method described in 3.2. It is therefore overestimated here.

3.4 Resistor loaded APD front end

If the photodiode is replaced with an APD with a gain M and an excess noise factor F, the above expression is modified in the following way:

$$n_{ph} \sim \frac{2SNR}{\eta}\sqrt{F\ n_{ph}I_2 + 4\pi\frac{C}{6.4aF}\frac{B_{RC}}{f_{data}}\frac{1}{M^2}I_2 + 4\pi\frac{C}{6.4aF}\frac{C}{C_{oa}}\frac{f_{data}}{f_T}\frac{1}{M^2}\gamma I_3} \tag{3.32}$$

Because of the intrinsic gain of the APD, the input referred noise for all noise sources is divided by M^2 (and thus the sensitivity associated with these sources is divided by M), while the shot noise is multiplied by F, along with the shot noise limit. All other considerations left aside, APDs are beneficial when the system is not shot noise limited, which is usually the case in non-coherent receivers.

3.5 Bipolar Phototransistor (BPT) front end

While the theory of operation of a phototransistor will be presented in chapter 5, the sensitivity is calculated here for comparison. If the collector is considered to feed a very low impedance so that its voltage stays constant, one end the capacitor C_μ can be considered to be connected to ground and the small circuit signal of a phototransistor (shown in figure

3.2) is mostly the same as the resistor loaded p-i-n case described earlier. The sources of noise are the base current shot noise $I_{N,B}$, the collector current shot noise $I_{N,C}$ and the signal shot noise $I_{N,S}$.

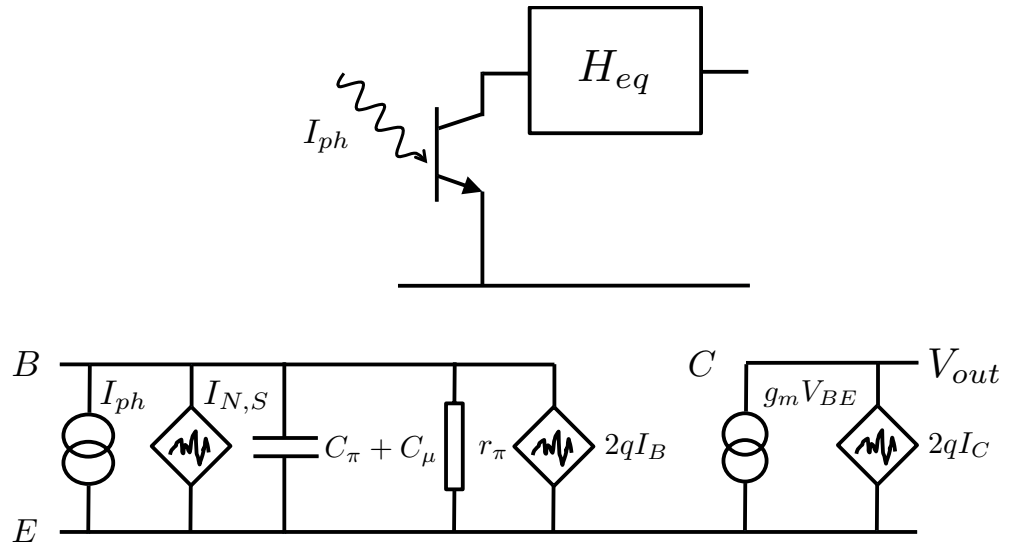

Figure 3.2: Bipolar phototransisitor schematic and small signal equivalent circuit

The different resistances and capacitances for a BPT are:

$$g_m = \frac{I_C}{V_{th}} \tag{3.33}$$

$$r_\pi = \frac{\beta}{g_m} \tag{3.34}$$

$$C = C_B + C_{diff} = C_{J,BE} + C_{J,BC} + \frac{\tau_F I_C}{V_{th}} \tag{3.35}$$

where τ_F is the transit time of the electrons through the device, and β is the DC gain of the device. The bandwidth of the front end is

$$B_{BPT} = \frac{1}{2\pi C r_\pi} \tag{3.36}$$

Using the same method as previously, the sensitivity of the front end is

$$n_{ph} \sim \frac{2SNR}{\eta}\sqrt{n_{ph}I_2 + 2\pi\frac{B_{BPT}}{f_{data}}\frac{C}{6.4aF}I_2 + 2\pi\frac{C}{C_{diff}}\frac{C}{6.4aF}\frac{f_{data}}{f_{t,m}}I_3} \tag{3.37}$$

with $f_{t,m} = 1/(2\pi\tau)$

3.6 Trans-impedance amplifier front end

The trans-impedance amplifier depicted in figure 3.3 is used as an example for the derivation of the input referred noise. This analysis can easily be generalized to other TIA configurations. The noise sources are the photon shot noise $I_{N,S}$, the resistor feedback noise $I_{N,R}$ and the transistor Johnson noise $I_{N,T}$.

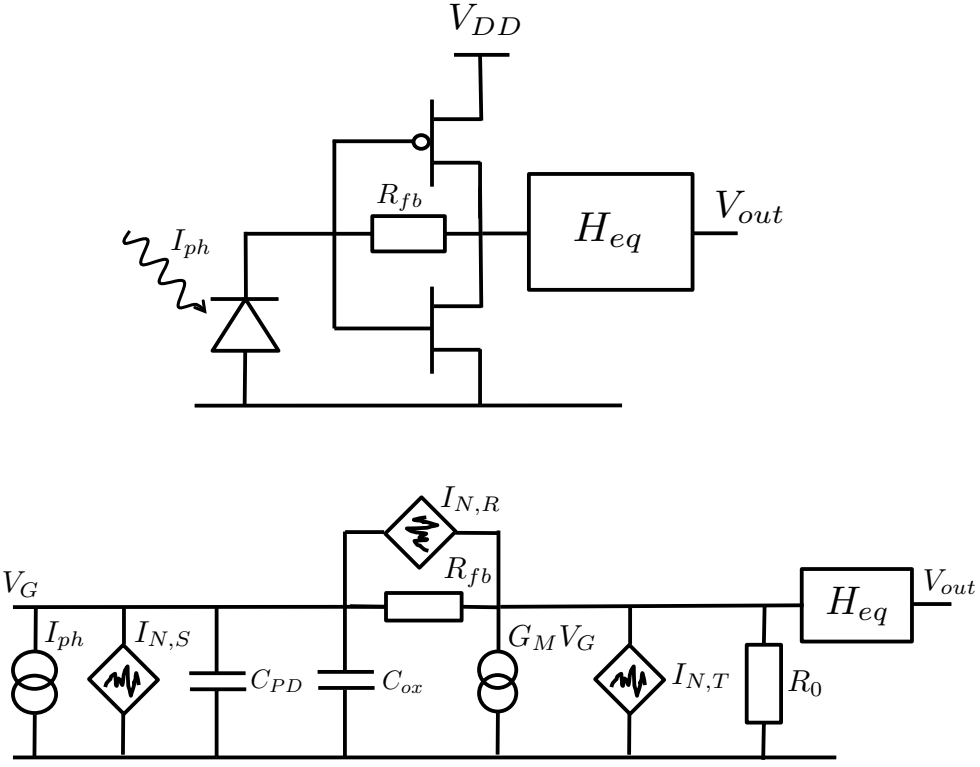

Figure 3.3: Transimpedance amplifier front end schematic and small signal equivalent circuit

With:

$$G_M = g_{m_P} + g_{m_N} \qquad \text{the combined transconductance of the FETs} \tag{3.38}$$

$$C = C_{PD} + C_{ox} = C_{PD} + \frac{G_M}{2\pi \tilde{f_T}} \qquad \text{the capacitance at the input} \tag{3.39}$$

$$Z_C = \frac{1}{jC\omega} \tag{3.40}$$

$$\tilde{f_T}^{-1} = (f_{T,PMOS}^{-1} + f_{T,NMOS}^{-1})/2 \quad \text{the effective unity gain frequency} \tag{3.41}$$

$$R_0 = R_{0,PMOS}||R_{0,NMOS} = \frac{G_{int}}{G_M} \quad \text{the output impedance of the FETs} \tag{3.42}$$

$$G_{int} \quad \text{the intrinsic DC gain of the transistors} \tag{3.43}$$

The transfer function of such a front end is

$$H(\omega) = \left[Z_C \frac{G_M R_{fb} - 1}{Z_C + R_{fb}}\right] \left[\frac{1}{R_0} + \frac{G_M Z_C + 1}{R_{fb} + Z_{PD}}\right]^{-1} \sim R_{fb} \frac{G_{int}}{G_{int}+1} \frac{1}{1 + j\frac{C(R_0+R_{fb})}{G_{int}+1}\omega} \tag{3.44}$$

where $G_M R_{fb} >> 1$ is assumed. And the well know formulas for the trans-impedance gain and bandwidth of a TIA is verified:

$$R_{TIA} = R_{fb} \frac{G_{int}}{G_{int}+1} \tag{3.45}$$

$$B_{TIA} = \frac{G_{int}+1}{2\pi C(R_0 + R_{fb})\omega} \tag{3.46}$$

The two noise sources of this front end are the transistor Johnson noise and the resistor Johnson noise. These can be expressed such as in equation 3.2 as:

$$V_{N_{out}} = \left[I_{N_R}\left[R_{fb}\frac{G_M Z_{PD}+1}{Z_{PD}+R_{fb}}\right] + I_{N_S}\right]\left[\frac{1}{R_0} + \frac{G_M Z_{PD}+1}{R_{fb}+Z_{PD}}\right]^{-1} + I_{N,S}H \tag{3.47}$$

Assuming, $G_M R - 1 \sim G_M R$, the sensitivity of the TIA front end is:

$$n_{ph} = \frac{SNR}{q\,\eta}\sqrt{2}\sqrt{I_{N,S}^2 T I_2 + I_{N,R}^2 T\left[I_2 + I_3\left(\frac{2\pi C}{T G_M}\right)^2\right] + I_{N,T}^2 T \frac{I_2 + I_3(RC2\pi/T)^2}{(G_M R)^2}} \tag{3.48}$$

Which can be approximated as

$$n_{ph} \sim \frac{2SNR}{\eta}\sqrt{n_{ph} I_2 + 4\pi\frac{C}{6.4aF}\frac{B_{TIA}}{f_{data}}\frac{1}{G_{int}}I_2 + 4\pi\frac{C}{6.4aF}\frac{C}{C_{ox}}\frac{f_{data}}{\tilde{f}_T}\gamma I_3} \tag{3.49}$$

3.7 Summary of noises, and transistor noise limit

The different front end noises are summarized in table 3.7.

Front end type	Minimum photons per bit
Resistor loaded p-i-n	$\frac{2SNR}{\eta}\sqrt{2n_{ph}I_2 + 4\pi\frac{C}{6.4aF}\frac{B_{RC}}{f_{data}}I_2 + 4\pi\frac{C}{6.4aF}\frac{C}{C_{ox}}\frac{f_{data}}{f_T}\gamma I_3}$
Resistor loaded APD	$\frac{2SNR}{\eta}\sqrt{2Fn_{ph}I_2 + 4\pi\frac{C}{6.4aF}\frac{B_{RC}}{f_{data}}\frac{1}{M^2}I_2 + 4\pi\frac{C}{6.4aF}\frac{C}{C_{ox}}\frac{f_{data}}{f_T}\frac{1}{M^2}\gamma I_3}$
Bipolar phototransistor	$\frac{2SNR}{\eta}\sqrt{n_{ph}I_2 + 2\pi\frac{C}{6.4aF}\frac{B_{BPT}}{f_{data}}I_2 + 2\pi\frac{C}{6.4aF}\frac{C}{C_{diff}}\frac{f_{data}}{f_{t,m}}I_3}$
TIA	$\frac{2SNR}{\eta}\sqrt{n_{ph}I_2 + 4\pi\frac{C}{6.4aF}\frac{B_{TIA}}{f_{data}}\frac{1}{G_{int}}I_2 + 4\pi\frac{C}{6.4aF}\frac{C}{C_{ox}}\frac{f_{data}}{f_T}\gamma I_3}$

In each of these cases, the first term under the root corresponds to the signal shot noise, the second term corresponds to the load resistor noise (is also sometimes referred to as the kTC noise) whereas the final term corresponds to the amplifying transistor noise.

Signal shot noise

The signal shot noise is naturally dependent on the input signal power, and is similar in all cases (except the APD). The quantum shot noise is the sensitivity limit that results from an ideal receiver with only signal shot noise, and as shown in 2, is roughly 10 photons per bit. In the calculations above, the symmetric noise postulate, along with the Gaussian noise approximation yield sensitivities substantially worse than this. Nonetheless even in this case, this crude approximation leads to a shot noise sensitivity of $4SNR^2 \sim 144$ photons per bit (~ 36 if corrected for noise asymmetry). For the resistor shot noise to be of the same order of magnitude, this would require the capacitance of the photodiode to be $> 50aF$. This points to the fact that in most cases we can ignore signal shot noise.

Resistor noise (aka kTC noise)

The resistor noise contribution is due to the random thermal fluctuation of charge in the capacitance on which the photocharge sits and is measured. It is therefore proportional to the capacitance, as well as it's coupling to a source of carriers. This is represented by the presence of the bandwidth of the system in the numerator.

The advantage of using a TIA becomes very clear with respect to the resistor noise contribution. Indeed the presence of feedback enables the use of a higher resistance value, therefore diminishing it's noise, by a factor of G_{int} compared to the simple resistor load.

Most of the time, the bandwidth is chosen to be roughly equivalent to the datarate, so that the ratio $\frac{B}{f_{data}} \sim 1$. Nevertheless this is a decision that is made only because it

is practical from a circuit design standpoint. Indeed it is possible to decrease the front end bandwidth and diminish the resistor noise. The reduction in bandwidth does not imply a reduction in gain-bandwidth, so the SNR is improved. Of course this requires equalization in the follow on stages to regain a flat band response. Practical limits to this technique are a reduction in dynamic range, since the low frequency components of the signal are amplified quite a bit more at the output of the front end. Nevertheless gains can be achieved, and the resistor noise can be brought lower than the transistor noise, as will be illustrated in chapter 4.

Transistor noise Limit

As discussed, the resistor noise can be arbitrarily diminished by lowering the bandwidth. This leaves us with the transistor noise. In the case of the TIA front end:

$$n_{ph,tran} = \frac{2SNR}{\eta}\sqrt{4\pi\frac{C}{6.4aF}\frac{C}{C_{ox}}\frac{f_{data}}{\tilde{f}_T}\gamma I_3} \tag{3.50}$$

We have $C = C_{PD} + C_{ox}$, and it can be easily shown that the optimal sizing in terms of noise for the transistor leads to $C_{ox} = C_{PD}$. This leaves us with a limit:

$$n_{ph,limit} = \frac{2SNR}{\eta}\sqrt{16\pi\frac{C_{PD}}{6.4aF}\frac{f_{data}}{\tilde{f}_T}\gamma I_3} \tag{3.51}$$

which we call the "transistor noise limit". It is roughly the same for all the proposed front end schemes (except naturally if an APD is used).

Chapter 4

Optical link modeling and performance analysis

In the previous chapter, the sensitivity of different front ends were calculated, and their sensitivity limits were explained. A full optical link is more complex and it's performance does not depend solely on the receiver front end sensitivity.

In order to guide receiver design, it is important to have a deeper understanding of the limits imposed by it's architecture. In this chapter a full optical link is analytically modeled from optical input to digital output, and subsequently the architecture of the receiver is optimized for full link energy consumption. The optimized receiver designs are then simulated in Cadence to validate the analytic models, and extract more accurate sensitivities and power consumption. The model is then used to predict the performance of optimal links for a variety of different technologies, and understand what the practical limits are.

The work in this chapter has been published in [11]

4.1 Link Model and optimization

A very general model for the optical link is considered, which enables us to perform optimizations on the links topology and to estimate of the optimal energy per bit which can be achieved at given datarates given the technology constraints. The philosophy of the model is depicted in figure 4.1 and as such: the receiver element is constructed with a transimpedance front end followed by N amplifications stages and terminated with a sampling unit composed of M individual samplers. The number of amplification stages N, the size of each stage, the number of sampling units M and the sizing of it's transistors constitute optimization variables. There is of course a variety of other receiver topologies or variations than the one suggested. The framework we describe next will be readily extendable to these topologies.

The energy consumed in the receiver can be computed from the bias currents and

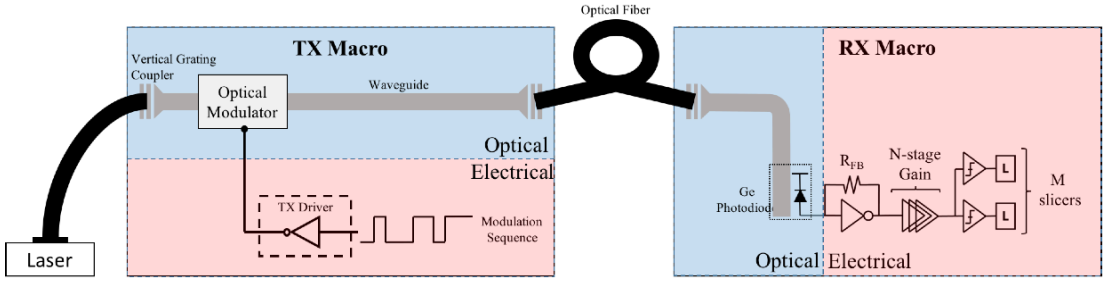

Figure 4.1: Optical link system overview

it's sensitivity is determined by two constraints: a noise constraint, and an output voltage constraint. Finally the energy consumed in the transmitter is calculated from the sensitivity of the receiver and the losses and inefficiencies along the path of the photon signal. The total energy is the sum of the receiver and transmitter energy, and is minimized with respect to the optimization variables at hand.

Receiver modeling

The receiver is modeled as illustrated in figure 4.1. The front end is a transimpedance amplifier (TIA) that converts the input photocurrent to a voltage signal, and is followed by N chained gain stages forming a linear amplifier (LA) to further amplify the signal . All these amplifiers are considered to be first order stages (except for the TIA which has two poles). The chaining of such stages causes the overall bandwidth to degrade. The bandwidth B_{chain} resultant from n_s first order stages of bandwidth f_S is [6]:

$$B_{chain} = f_S \frac{0.9}{\sqrt{n_s + 1}} \tag{4.1}$$

Since the total bandwidth should be at least $0.7 \times f_{data}$ to minimize ISI, this implies that the bandwidth f_S of each stage must be

$$f_S > 0.7 f_{data} \frac{\sqrt{(N+2)+1}}{0.9} \tag{4.2}$$

in order to satisfy this constraint, where both poles o the TIA have been taken into account.

Gain-Bandwidth product

While the unity current gain-bandwidth of a technology is f_T, the actual gain bandwidth that is achieved in an individual gain stages that is loaded by it's replica will be lower due to various parasitics and non idealities. Additionally, different gain stage topologies will

yield different GBWs, for example inductive peaking is a popular way of enhancing the bandwidth and will yield a higher GBW than simple resistively loaded stages. Therefore a parameter α is used which describes what fraction of f_t is achieved by each individual gain stages. The GBW of a replica-loaded stage therefore $f_a = \alpha f_T$.

Linear amplifier

Every stage in the linear amplifier is defined by its input transistor gate width $W_{gate,i}$ (where "i" denotes its position in the amplifier chain), which then also defines its transconductance $g_{m,i}$, gate capacitance $C_{ox,i}$ and bias current $I_{d,i}$. To simplify the problem, it is assumed that f_T, C_{ox}, and I_d are simply proportional to W_{gate}, which implies that the biasing for each transistor is relatively similar: a reasonable assumption to first order. The GBW of each stage depends on the capacitance seen at the output, and in the case of simple resistively loaded stages: $GBW_i = g_{m,i}/(C_{out,i} + C_{in,i+1})$. Additionally $\beta = C_{out}/C_{in}$ is defined as the ratio of output to input capacitance of a gain stage. Similarly to α, β is dependent on stage topology. A table of these parameters is given and derived in the appendix for different topologies of gain stages. Therefore

$$f_a = g_m/((1+\beta)C_{in}) \tag{4.3}$$

and the GBW of every stage can be derived as:

$$GBW_i = \frac{g_{m,i}}{C_{out,i} + C_{in,i+1}} = f_a \frac{1+\beta}{\beta + \frac{W_{gate,i+1}}{W_{gate,i}}} \tag{4.4}$$

As mentioned earlier, each gain stage must also have a 3-dB bandwidth of f_S, so that the DC gain of stage i in the linear amplifier is:

$$G_{DC,N} = \frac{f_a}{f_S} \frac{1+\beta}{\beta + \frac{W_{gate,i+1}}{W_{gate,i}}} \tag{4.5}$$

The maximum gain is capped by the intrinsic gain of the devices: $g_m r_0$. For the last stage, the capacitance driven is the sampler's input capacitance C_{SA}. Finally the power consumed by each stage is $V_{DD} I_{bias}$, where $I_{bias,i} = g_{m,i} V_{ov}$, where V_{ov} is the stage overdrive voltage (considered to be the same for every stage).

Transimpedance amplifier

The transimpedance amplifier is composed of a gain stage similar to those in the LA, with a feedback resistor chosen in order to meet the bandwidth requirement per stage f_S. The

open loop gain is calculated the same way as described previously for the LA stages, and the feedback resistor is therefore set to:

$$R_{FB} = \frac{G_{DC,TIA}}{2\pi f_S(C_{PD} + C_{in,TIA})} \tag{4.6}$$

where C_{PD} is the photo-detector parasitic capacitance including the interconnect between the photo-detector and the TIA, and $C_{in,TIA}$ is the TIA input capacitance. The two poles resulting from the TIA designed in this fashion are not real, and the damping factor is $\zeta = \frac{1}{2}\frac{2+G_{DC,TIA}}{1+G_{DC,TIA}}$ bounded as $0.5 < \zeta < 1$, implying the bandwidth is marginally greater than if the poles were real. This means that equation 4.2 slightly overestimates the required bandwidth per stage. To first order this is an acceptable approximation.
The total transimpedance gain of the TIA and the LA is

$$R_{tot} = \frac{R_{FB}\ G_{DC,TIA}}{1 + G_{DC,TIA}} \prod_{i=1}^{N} G_{DC,i} \tag{4.7}$$

Sampler

The modeled sampling stage is made of M interleaved StrongArm samplers (also referred to as Sense Amplifiers (SA)), that evaluate the bits sequentially. This means each individual strongARM has a cycle $M \times T_{bit}$ long. Half of this period is dedicated to the resetting of the sampler, while the other half is dedicated to the integration and regeneration of the bit. The schematic of an individual sampler is depicted in figure 4.2 and the waveforms associated with it are shown in figure 4.3. The integration period lasts while the input pair discharges nodes P,Q,X and Y, and lasts until nodes X and Y reach $V_{DD} - V_{th,P}$ which dictates when the cross coupled pair turns on and the regeneration period starts [12] ($V_{th,P}$ is the threshold voltage of the PMOS). Figure 4.3 shows a StrongArm's transient characteristics with the three main regimes of operations highlighted. The regeneration gain is generated by a cross coupled pair forming a latch, is exponential with time, and brings the output signal to logic levels.

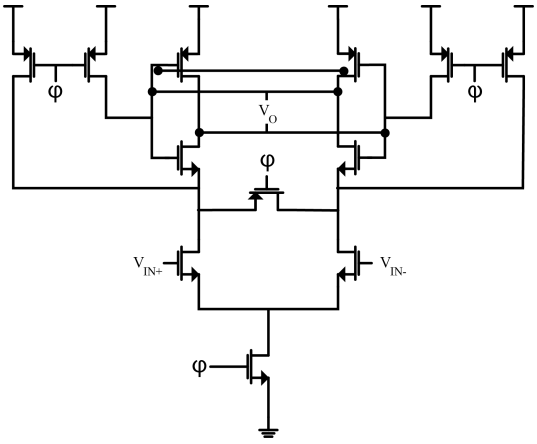

Figure 4.2: StrongArm Sampler Schematic

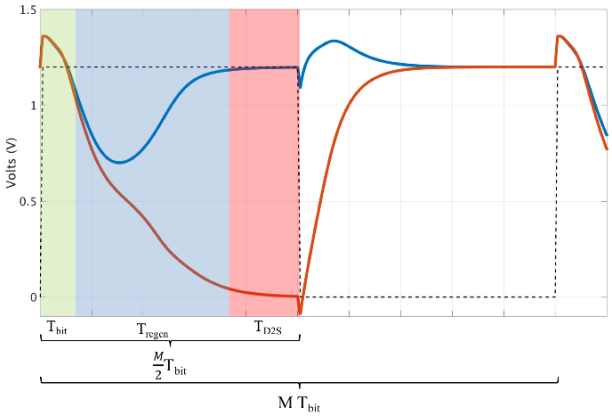

Figure 4.3: Sampler timing evaluation breakdown

The optimization variables available are the common mode voltages at the input, the gate widths of the input transistors, and the gate widths for the cross coupled pair transistors. These define the length of the integration period (which must stay under T_{bit}), the integration gain and the regeneration gain. The sampler then drives a dynamic to static (D2S) converter stage which is simply characterized as a load capacitance to the sampler $C_{in,D2S}$ [13]. The D2S needs a certain amount of time $T_{D2S} \sim \frac{2}{f_T}$ to latch, which is taken out of the total evaluation time. The derivation of the integration time and sampler gain are derived in appendix A. Approximations are nevertheless give here:

$$T_{int} \sim \frac{V_{TH}(2C_{PQ} + C_{XY})}{g_{m,1}(V_{CM} - V_{TH})} \tag{4.8}$$

$$G_{int} \sim \frac{V_{TH}}{V_{CM} - V_{TH}} \frac{C_{PQ} + C_{XY}}{C_{XY} - C_{PQ}} \tag{4.9}$$

$$T_{SA} = M/2 \times T_{bit} - T_{D2S} \tag{4.10}$$

$$G_{SA} \sim G_{int} \exp(\frac{M/2 \times T_{bit} - T_{int} - T_{D2S}}{\tau_{reg}}) \tag{4.11}$$

$$\tau_{reg} = \frac{g_{m,3} + g_{m,5}}{C_{in,D2S} + C_{out,SA}}; \tag{4.12}$$

Where V_{TH} is the absolute value of the threshold voltages. Finally the input capacitance of the SA seen by the front end is given by $M \times C_{ox,SA}$. The fanout M is detrimental to the gain of the front end, and, as will be shown, can be amortized by using switches that connect only one sampler at a time to the output of the sampler. In this case, the input capacitance seen by the sampler is approximately $C_{ox,SA}$ neglecting junction capacitance effects of the sampling switches and the RC time associated with them. This assumption holds true for reasonable number of samplers:

$$M < \frac{f_T}{f_{data}} \frac{C_{ox}}{C_{gd}} \tag{4.13}$$

Indeed the size of the transistor serving as a switch can be made substantially smaller than the input cap of the SA, by a factor $\sim \frac{f_T}{f_{data}}$ to minimize it's effect on the circuit bandwidth, and the only capacitance it presents to the circuit is it's gate-drain capacitance C_{gd}, justifying equation 4.13.

The energy consumed by the sampler comes from the charging and discharging of all it's capacitances at each cycle, as well as the dynamic power burned by the cross coupled inverter during the latching process:

$$E_{samp} = E_{Cap} + E_{latch} \tag{4.14}$$

$$E_{Cap} = C_{SA} V_{DD}^2 \tag{4.15}$$

$$E_{latch} \sim (g_{m,3} + g_{m,5})(\frac{V_{DD}}{2} - V_{TH}) V_{DD} (T_{SA} - T_{int}) \tag{4.16}$$

where C_{SA} comprises all the capacitances that will have to be charged to V_{DD} during the reset period.

Sensitivity calculation

The sensitivity is the necessary amount of photon current needed in order for the system to function properly at a certain BER. It can be separated into two parts: the swing re-

quirement, and the circuit noise requirement. The final sensitivity is the sum of the two.

Swing Based Sensitivity Requirement

The swing requirement represents the signal needed to ensure that the differential voltage at the output of the sampler reaches V_{DD} and is calculated from the sampler gain, the TIA gain and the LA gain:

$$I_{req,swing} = 2\frac{V_{DD}}{R_{tot}G_{SA}} \tag{4.17}$$

The factor of 2 comes from the fact that the signal is only half the actual photon current magnitude for a optical ONE.

Noise Based Sensitivity Requirement

The noise requirement necessitates the calculation of the input referred noise generated by the amplification circuit. These include the feedback resistor thermal noise, the Johnson noise from the TIA's transistors, and the transistor noise from the follow on transistors as well as the noise from the samplers. All the relevant derivations were performed in chapter 3, but are listed here for convenience (with the different Personick integrals set to 1). The follow-on stage noises are estimated using approximations consistent with literature [14]. The photon shot noise (or PD shot noise) is neglected as it is always much lower than the circuit noise sources for incoherent detection systems (roughly one order of magnitude). Indeed for a BER of 10^{-12}, the limit that would be imposed by photon shot noise is 27 photons per bit during a ONE (also known as the quantum limit), which is a current of 44 nano-Amps at 10 Gbps.

$$I^2_{noise,in,R_{fb}} = \frac{4k_b\theta}{T_{bit}R_{fb}} \tag{4.18}$$

$$I^2_{noise,in,TIA} = \frac{16\pi^2 k_b\theta\gamma(C_{PD}+C_{in,TIA})^2}{g_{m,TIA}T^3_{bit}} \tag{4.19}$$

$$I^2_{noise,i} = \frac{4k\theta\gamma}{g_{m,i}[T_{bit}R_{fb}\prod G_{DC,j}]^2} \tag{4.20}$$

$$V^2_{noise,SA} = \frac{8k_b\theta\gamma}{t_2 g_{m1}} + \frac{8k_b\theta\gamma g_{m,3}}{t_{12}g^2_{m1}} + \frac{2k\theta}{C_{out,SA}G^2_{sample}} \tag{4.21}$$

Finally the sensitivity is calculated using a SNR of 7 in order to achieve a bit error rate of 10^{-12}.

$$I_{req,noise} = 2SNR\ I_{noise,input} \tag{4.22}$$

The total photon current requirement at the input of the photodiode is $I_{req,input} = I_{req,noise} + I_{req,swing}$.

Energy per bit

The total energy per bit that is consumed by the link is the sum of the energy burned in the transmitter and the receiver

$$E_{bit} = E_{RX} + E_{TX} \tag{4.23}$$

$$E_{RX} = T_{bit} V_{DD} \sum I_{bias} + E_{samp} \tag{4.24}$$

$$E_{TX} = T_{bit} V_{TX} (I_{req,noise} + I_{req,swing}) + E_{mod} \tag{4.25}$$

E_{TX} includes laser energy and modulator energy E_{mod}, where V_{TX} represents the energy cost of photons at the receiver: it encompasses all the efficiencies η encountered from the generation of photons to their absorption into useful photo-current in the receiver photodiode, such as the laser wall plug efficiency, coupler inefficiencies, waveguide losses, modulator loss, photodiode quantum efficiency, etc...

$$V_{TX} = \frac{h\nu}{q} \frac{1}{\eta} \tag{4.26}$$

$$\eta = \prod \eta_{system} \tag{4.27}$$

Model inputs and optimization variables

The model described enables the rapid prediction of the performance of a given optical receiver characterized by the number of amplification and sampling stages, the technology available, and the size of the transistors involved. These different parameters can therefore also be optimized in order to reach minimal total link energy. The optimization variables and model parameters are described in table 4.1, and the optimized links are presented in figures 4.5, 4.6 and table 4.2 .

Model purpose and limitations

The goal of the model is to accurately encompass all the most important effects and limits that fundamentally constrain the performance of an optical link. Naturally, no model can include all practical limitations, such as systemic and random transistor mismatches, kickback, jitter, layout imperfections, etc. Additionally exotic amplifications schemes such as higher order stages, or multiple interleaving schemes are not included. While these considerations are important in practical circuit design, we consider them to be nuances when studying the trends and do not drastically affect the general conclusions we derive from this model. It is nonetheless precise enough to provide optimal transistor sizing and accurate sensitivity predictions leading to functional circuits as shown in section IV.

Table 4.1: Model inputs and optimization variables

Model inputs	Variable description	65nm heterogeneous integration
f_t	Technology unit current gain frequency	150 GHz
C_{PD}	Photodiode capacitance	20 fF
f_{data}	Datarate	1 Gbps to 50 Gbps
α	Fraction of f_t for self loaded stage GBW	0.29 (standard $g_m R_L$ stages) 0.4 (cascode stages)
β	Fraction of input to output cap of a gain stage	0.67
$C_{out,D2S}$	D2S input capacitance	3 fF
t_{D2S}	D2S latching time requirement	25ps
V_{DD}	Supply voltage	1.6V
V_{TX}	Voltage cost of photons	580 V
$(g_m r_0)_{max}$	Maximum voltage gain per stage	4
V_{ov}	Overdrive voltage	0.3 V
γ	noise factor	2/3
E_{mod}	Modulator energy per bit	0
Optimization variables		**Bounds**
N	Number of amplification stages	0 to 4
M	Number of samplers (M-DR)	1 to 64, in powers of 2
$W_{gate,TIA}$	Input transistor gate width for the TIA	$>$ 150nm
$W_{gate,1,...,N}$	Input transistor gate width for each stage	$>$ 150nm
$W_{gate,in,SA}$	Input transistor gate width for sampler	$>$ 150nm
$W_{gate,CC,SA}$	Transistor gate width for cross coupled pair	$>$ 150nm
V_{CM}	Common mode voltage at SA input	0.8V$<$ ... $<$1.4V $=V_{DD}/2$ if N=0

4.2 Model results for 65nm technology with heterogeneously integrated photonics

Technology Overview

The model is used to output optimal designs for a technology made of 65nm CMOS, with a heterogeneously integrated photonics, as depicted in figure 4.4. [15]. To reduce the capacitance between the CMOS and photonic wafers, the technology utilizes through-oxide-VIAs (TOVs) with a lumped capacitance of 3fF per TOV. The major technology parameters are listed in Table 4.1.

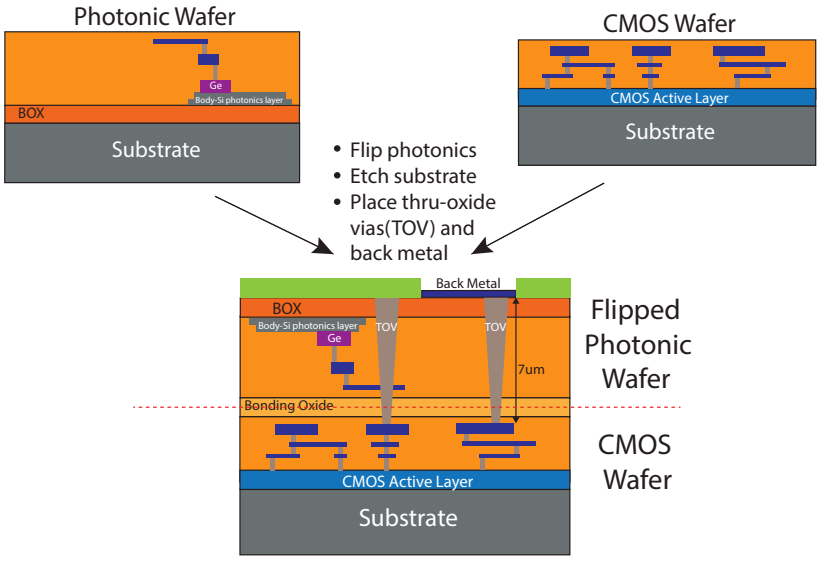

Figure 4.4: Heterogeneous integration platform schematic, from [15]

Using Figure 4.1 as reference, light from the laser source experiences multiple sources of loss before reaching the photodiode on the receiver side. Firstly, the laser source itself is assumed to have a wall-plug efficiency of 20%. The three vertical grating couplers, which measured 6.5dB/coupler of loss, are also in the critical path of the signal. The germanium photodiode has a measured responsivity of 0.8 A/W [16]. No waveguide loss is assumed, however this can be easily implemented. The above path losses translate to an overall photon energy cost, V_{TX}, of 580. The modulator energy in this platform is 20fJ/bit and will therefore be neglected.

Single sampler case (M=1)

The results of the optimization for the optimal performance of the link are plotted in figure 4.5. The laser energy to accommodate the noise and swing requirements are respectively the quantities described in equation 4.25. Two clear regimes are visible: the "Noise limited regime" at low datarates, where the sensitivity of the receiver is constrained by the noise, and the "Swing limited regime" at high datarates, where the sensitivity is dominated by the output swing requirement ($V_{out} = V_{DD}$). The regeneration gain of the sampler is exponential with time, so it is natural that at higher datarates it drops significantly. While the LA can compensate for this drop in gain by increasing it's number of stages (and this happens at $\sim$ 8.5GHz for case I), there is a limit to the amount of aggregate gain achievable by chaining amplifiers due to the bandwidth requirement, as described in equation 4.2.

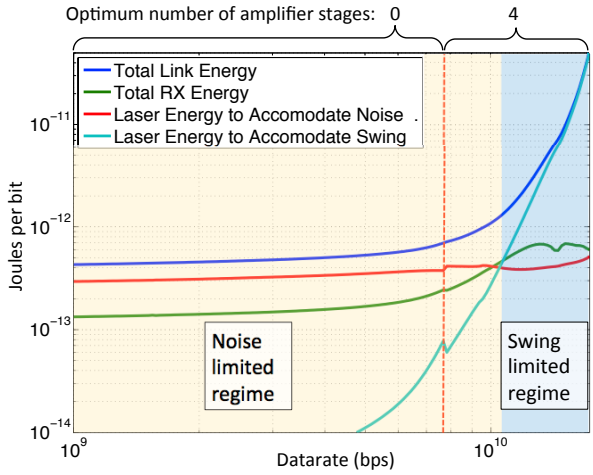

Figure 4.5: Optimal energy per bit versus datarate for optimal topologies for the 65nm heterogeneously integrated platform. Only one slicer his allowed in his case

The justification for adding multiple slicers is now obvious: this relaxes the condition on the regeneration time being less than the bit duration, and can push the swing limited regime to much higher datarates.

Multiple slicer case ($M \geq 1$)

The results of the optimization when the number of samplers is not constrained to 1 is plotted on figure 4.6. There is no longer a "Swing limited regime", since the optimal topologies have several samplers in order to benefit from much higher regeneration time and gain. While the energy per bit is greatly reduced at higher datarates, eventually the sampler noise starts to dominate. This comes about because as the datarate goes up, the bandwidth requirement on the LA reduces the possible achievable gains. Additionally, adding several samplers increases the fan-out of the LA by a factor M. Eventually the fanout becomes greater than the gain of the stages before the samplers, so that the noise coming from the samplers becomes greater than the front end noise. Therefore a front end noise limited regime and a sampler noise limited regime are observed, which is different from the sampler swing limited regime discussed in the SDR case.

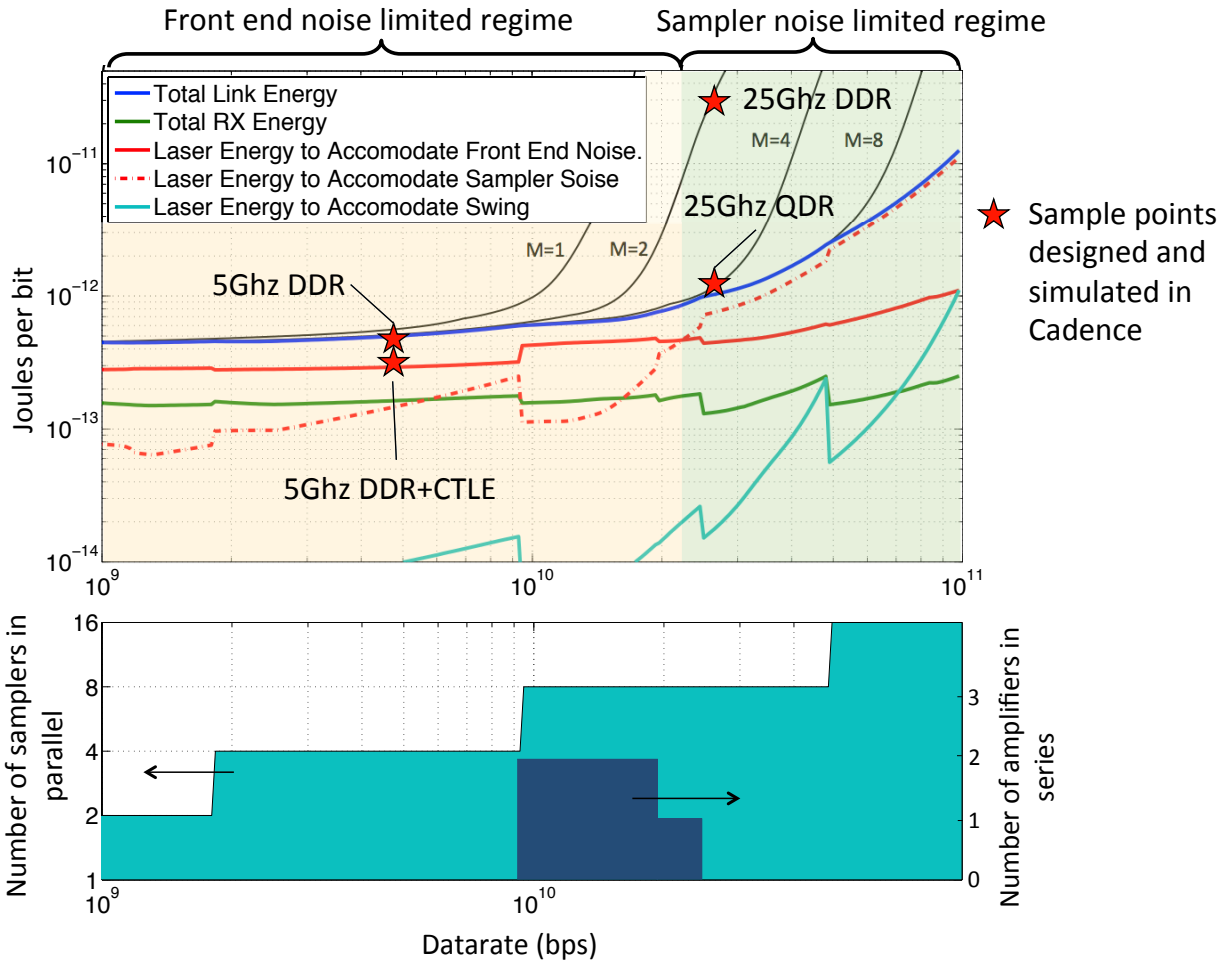

Figure 4.6: Optimal energy per bit versus datarate for optimal topologies with parameters of case 1, with the possibility of multiple slicers

4.3 Schematic designs of model results

5Gbps Optical Receiver

To highlight performance in the noise-limited regime, the schematic design of an optimized receiver topology operating at 5Gbps, with no active equalization, running off of a 1.2V supply is introduced. Figure 4.7 shows the overall topology of the front-end pre-amplifier and slicers. While the number of slicers and samplers does not match the optimal values of figure 4.6, these values were chosen because they yielded performance within a few percent of the optimum, and were easier to implement. Nonetheless the transistors sizings were still produced by the algorithm.

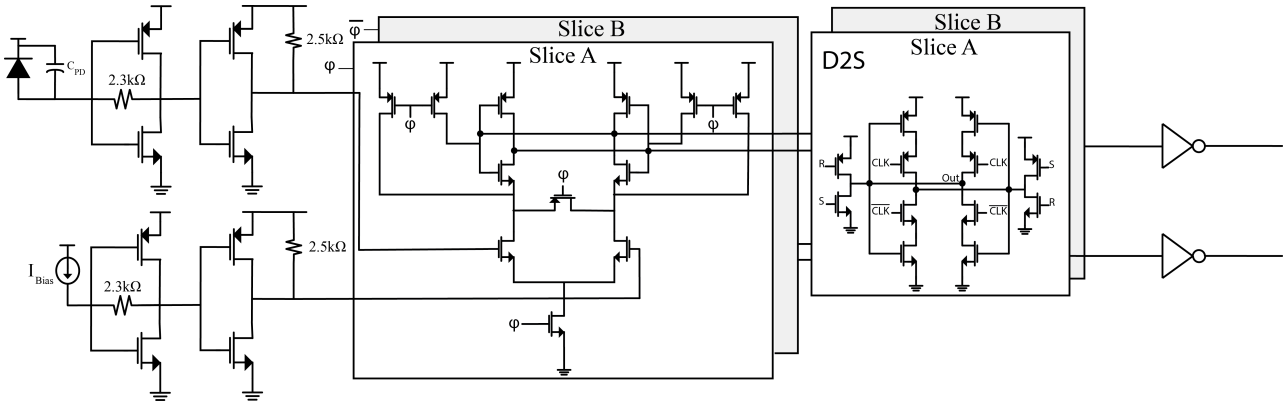

Figure 4.7: 5 Gbps Model-Predicted Receiver Topology

Design Overview

The photodiode, with a total capacitance C_{PD} of 20fF, inputs into a TIA amplifier with a feedback resistance, R_{FB}, valued at $2.3k\Omega$. The output of this stage enters a single pre-amplifier gain stage with a gain of 2 before entering the optimized, dual-data-rate (DDR) triggered StrongArm Sense Amplifiers and follow-on dynamic-to-static converters. The sense amplifiers and dynamic-to-static converters are triggered on clock and clockB (Φ and Φ_B), which each operate at half the data rate or 2.5GHz. The sampler transistor sizes as well as the front-end sizings are optimized using the algorithm. Additionally, the biasing at each stage is also dictated by the algorithm. More specifically, the common-mode voltage at the input of the samplers was selected to be 850mV while the constrained common-mode voltages at the TIA's input and output were set at $V_{DD}/2$ or 600mV. The output of the slicers, which are effectively a 1-to-2 deserialized version of the input data sequence, was verified in simulation.

Simulation Results

The above design has been implemented at the simulation level and its performance was verified with respect to the values predicted from the model. Table 4.2 summarizes specifications for the model and simulated results. The optimized circuit had an overall front end gain of $5.1k\Omega$ and from the StrongArm sampler's standpoint, the minimum required swing at the input (neglecting noise) to resolve successfully at 5Gbps, or 200ns of evaluation time per sampler, was measured to be 6mV. This translates to a 1.2μA receiver sensitivity due to the swing requirement of the sampler. From a noise perspective, the total input-referred noise contribution from the front-end is 0.04μA (1σ). Thus, the total simulated input sensitivity is 3.8μA. The total energy per bit for the full RX block is 280 fJ/bit, with the front-end consuming 115 fJ/bit and the samplers plus D2S consuming 165 fJ/bit total. The front-end E/b in this case takes into account the the dummy front-end as well. From an overall link perspective, the energy breakdown in the laser and TX macro are 392fJ/bit.

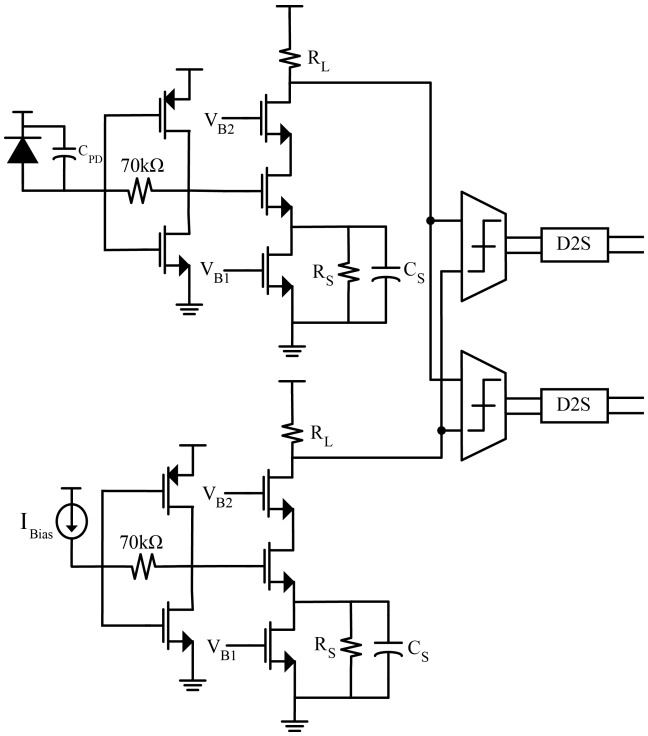

Figure 4.8: 5Gbps Model-Predicted Receiver Topology with Active-CTLE

Active-CTLE Enhanced 5Gbps Optical Receiver

The algorithm is now used to design a continuous-time linear equalizer (CTLE)-based optical receiver front-end. The purpose of this design is to show that it is possible to reduce the noise contribution of the feedback resistor in the front end, as will be discussed in section 4.4. The circuit is nevertheless presented here for consistency. The full schematic is shown in Figure 4.8.

Design Overview

The CTLE-based pre-amplifier allows for the preceding TIA stage's R_{FB} to increase drastically, from $2.3k\Omega$ to $70k\Omega$. This enhances noise performance while keeping the overall gain-bandwidth the same. The new low-pass pole of the TIA front-end is then compensated with the peaking of the CTLE block, which adds a *zero* in the transfer function from the degenerated R_S and C_S (see Figure 4.11). The *zero* location is chosen to be the dominant pole-location of the TIA, thereby enhancing the overall bandwidth to the target specification for operation at 5Gbps. The two poles of the CTLE are set to the same frequency in order to maximize the effective gain-bandwidth of the stage [17].

Results

The results are summarized in Table 4.2. The overall receiver gain and bandwidth of the CTLE are approximately that of the standard RX topology, at 5460 kΩ and 5.3GHz, respectively. The CTLE-based front-end consumes 250 fJ/bit with the samplers consuming 141 fJ/bit. This yields an overall RX E/b of 391 fJ/bit. The main advantage in using a CTLE-based scheme comes from the input referred noise sensitivity. Here, we observe 0.2μA input sensitivity whereas the standard RX topology had almost double that. In the CTLE topology, the feedback resistor contributes only 15% of the total front-end noise, whereas the standard RX topology's contribution is almost 50%.

DDR 25Gbps Optical Receiver

Design Overview

To better characterize the universality of the model, an optimized optical receiver design operating in the sampler swing-limited regime is presented. The circuit operates at V_{DD} of 1.6V in order to allow for enough voltage headroom to utilize cascode-amplifiers as the basis design for the VA stages, which have an α of 0.4 as opposed to the standard amplifiers which have an α of 0.29. The StrongArm topology for the samplers as well as the topology of the D2S converters is retained. In this design, the system operates as DDR (2 samplers) to show the importance of relaxed timing margin on the sampler's evaluation period.

Under these constraints, the model-predicted topology is shown in Figure 4.9. All front-end FETs, resistances, and sampler FETs, are all sized based on the constraints presented by the algorithm. In the DDR case, M=2

Results

To avoid choking the bandwidth at the input node of the TIA itself, the model-predicted TIA feedback resistance was 530Ω. This translates to an overall gain of 770Ω in the two-stage front-end and an overall bandwidth of 18.8 GHz, which meets our programmed target specification of 0.7*25GHz, or 17.5GHz. At this data rate, the sampler required a minimum swing of 165mV with a common mode of 840mV. The overall swing-based sensitivity is therefore 280μA. The rationale for this high sensitivity is as follows: because the system was operating within the sampler-swing dominated regime and with a fixed number of samplers for DDR, the algorithm would resort to increasing the laser power to meet the sensitivity requirement of the sampler instead of adding further amplification stage, which is not possible due to the bandwidth degradation penalty. In this regime the input referred sensitivity due to noise is, as expected, very small compared to the sensitivity requirement due to swing. The overall power breakdown shows 395fJ/bit consumed in the front-end and 153fJ/bit consumed in the samplers. The total RX E/b is 550fJ/bit.

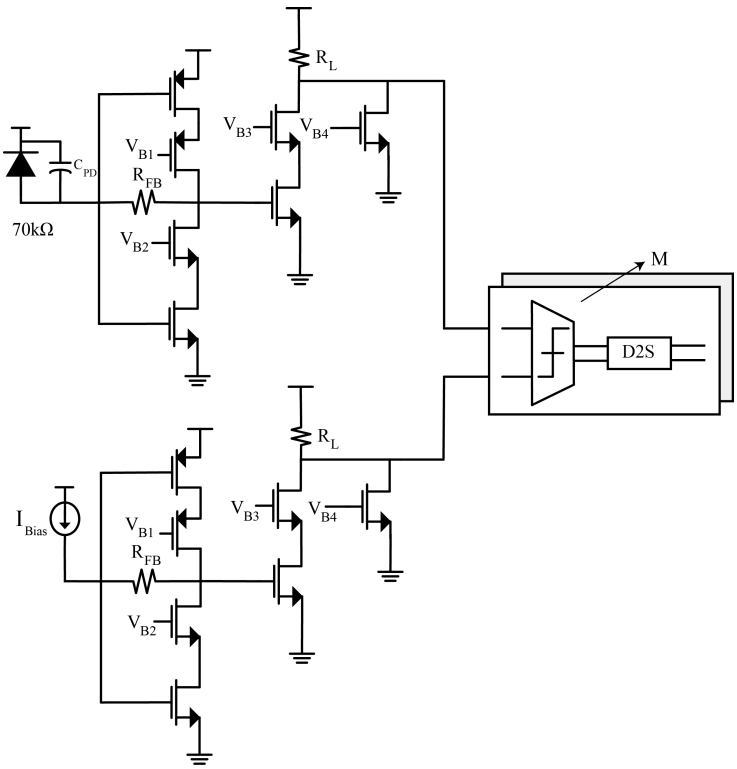

Figure 4.9: 25 Gbps Model-Predicted Receiver Topology

QDR 25Gbps Optical Receiver

Design Overview

In the subsequent analysis, we retain the same technology parameters as in the previous section. However, now, we present a quadrature-data-rate (QDR), M=4 from Figure 4.9, operation of the receiver, wherein four samplers are utilized to parse the amplified photodiode signal. Once again, the design of the front-end as well as samplers is fully predicted with the tool optimizing for the added capacitive load factor on the final stage of the pre-amplifiers. In using four phases, the timing evaluation requirements of the samplers are alleviated by doubling the allocated time for sampling and reset phases, while adding clocking overhead in the form of quadrature phase generation. In the context of links, this drastically improves efficiency and extends the crossover point of the noise-limited and sampler-limited regimes to past 25Gbps, as seen in Figure 4.6.

Results

The QDR receiver performed on par with the DDR in power, gain, and bandwidth metrics. However, from a swing sensitivity standpoint, the QDR receiver performed an order

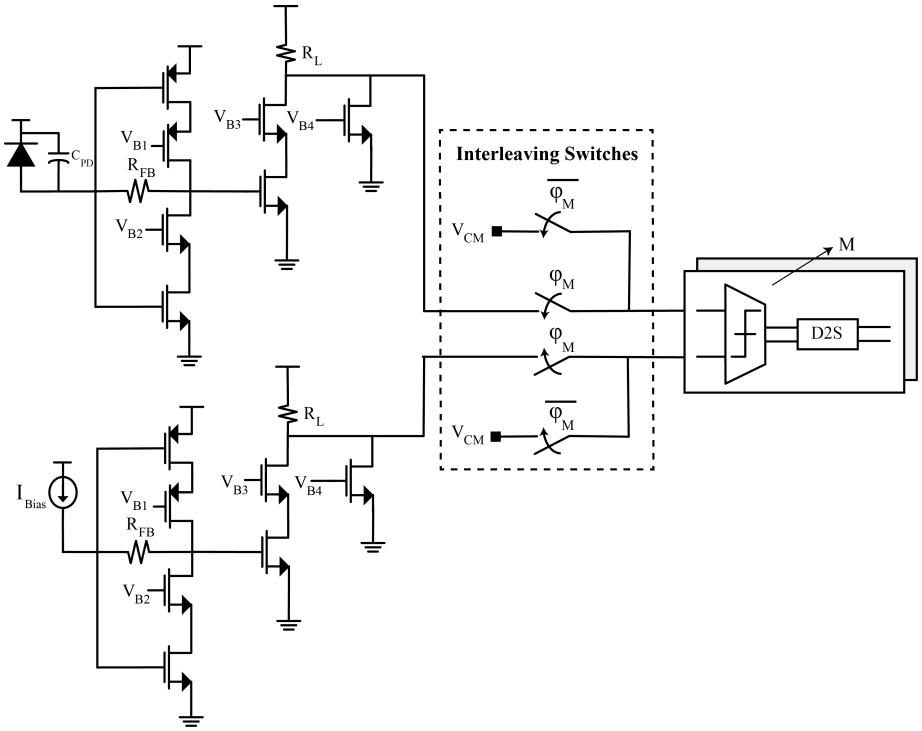

Figure 4.10: Switching Time-Interleaved 25 Gbps QDR Receiver

of magnitude better. The simulations yielded a swing sensitivity of under 5μA, with a front-end gain of 760Ω and 20.5GHz net bandwidth. The four samplers and D2S's were consuming 153fJ/bit while the front-end was consuming 395fJ/bit for a total 550fJ/bit being burned on the receiver end. The input referred noise sensitivity for the receiver was 1.8μA, now mostly dominated by the sampler noise.

Because of this ultra-low sensitivity, even though the RX total power stayed approximately the same for the DDR and QDR cases, the required laser power was substantially reduced, as shown in Table 4.2.

Switched QDR 25Gbps Optical Receiver

Design Overview

To alleviate the sampling noise contribution of the StrongArm sense amplifiers, a time-interleaved switching topology was implemented, reducing the load on the VA and allowing it to provide more gain for a given bandwidth constraint. The schematic is shown in Figure 4.10. By placing a track and hold circuit prior to the sampler array, not only does the sense amplifier input load capacitance diminish, but the potential effects of kickback from other sampler clocks is also theoretically reduced. The receiver topology and design process is similar to the QDR receiver in Figure 4.9. All transistor sizings are model-predicted

with the biggest difference being in how the $C_{ox,SA}$ capacitance scales. $C_{ox,SA}$ now goes up linearly with sampler input FET size and is completely independent of the slicing count, M, as detailed in Section 2. For the purposes of this study, the non-idealities of these sampling switches (i.e. finite junction capacitance) were not taken into account within this study. However, the simulation results reflect performance with these non-idealities in place, and we see no significant difference between the predicted and simulated specifications. This is because the sampler count, M, is kept to a reasonable value according to equation 4.13.

Results

The results in 4.2 for the 25Gbps Switching QDR receiver show similar performance to the non-switching. However, the total noise sensitivity is reduced by 10% on account of the sampler noise contribution reducing, while the noise from the front-end stays relatively constant. The sensitivity required to overcome the sampler swing is also relatively constant, with small adjustments made to the input sampler FETs on account of the switching.

Table 4.2: Performance Comparison of Model-Predicted and Schematic-Simulated Optical Receivers

Specification	5Gbps Standard RX		5Gbps CTLE RX		25Gbps DDR		25Gbps QDR		25Gbps Switching QDR	
	Modeled	**Designed**	**Modeled**	**Designed**	**Modeled**	**Designed**	**Modeled**	**Designed**	**Modeled**	**Designed**
Total Front-end Gain (Ω)	15500	20200	15500	15500	636	770	1055	760	1750	1840
Total Front-end BW (GHz)	3.5	3.4	3.5	3.4	17.5	18.8	17.5	20.5	17.5	17.8
Total AFE E/b (fJ/bit)	79	52	79	91	340	300	260	395	134	135
Total Sampler E/b (fJ/bit)	32	35	32	37	110	97	93	153	88	91
Total RX Sensitivity (μA)	3.37	3.8	2.95	3.2	356	288.2	33.9	86.2	28	77.4
SA Input Common Mode (mV)	500	500	500	530	850	840	850	810	850	950
SA Minimum Swing (mV)	0.2	6	0.2	5	210	165	1.6	4	1.5	11
RX Swing Sensitivity (μA)	0.01	0.8	0.01	2.1	326	214	1.7	5	0.9	6
14σ RX Noise Sensitivity (μA)	3.4	2.9	2.9	1.1	30.8	74.3	32.2	81.2	26.6	71.4
Link RX E/b (fJ/bit)	110	88	110	128	440	400	225	550	221	227
Link Laser E/b (fJ/bit)*	392	–	337	–	16000	–	840	–	655	–

4.4 Sensitivity and energy limits

While the model enables the choice of optimal transistor sizings and system link efficiencies, it does not immediately provide a deep understanding of the different limits experienced by such a system. In this section these limits are derived.

As shown earlier, it is possible to alleviate the swing requirement by using an appropriate amount of interleaved samplers. In a similar way, if a sample and hold method is used as in Section 4.3 to negate the effect of fanout, the dominant noise source comes from the very front end, which is therefor what this section will focus on.

Front end noise limit

The noise in the front end is dominated by the first amplification stage, which is the TIA in this case. The two major sources of noise have been given in Equations 4.18 and 4.19, and their input referred noise current is given in Equations 4.28 and 4.29.

$$I_{n,R}^2 = (qf_{data})^2 8\pi \frac{C_{PD} + C_{in,TIA}}{6.4aF} \frac{f_{TIA}}{f_{data}} \frac{1}{g_m r0} \tag{4.28}$$

$$I_{n,amp}^2 = (qf_{data})^2 8\pi \frac{(C_{PD} + C_{in,TIA})^2}{6.4aF \times C_{in,TIA}} \frac{f_{data}}{f_T} \frac{\gamma}{\alpha(1+\beta)} \tag{4.29}$$

$$\tag{4.30}$$

Where f_{TIA} is the bandwidth of the TIA, and $6.4aF = q/V_{th}$ where V_{th} is the thermal noise voltage. The optimal $C_{in,TIA}$ that minimizes the sum of both noises is somewhere between 0 and C_{PD}.

Nevertheless, the feedback resistor noise can be overcome to some extent by increasing the value of the feedback resistor, and compensating for the bandwidth degradation by including equalization such as a CTLE stage, as we show in the example circuit in Figure 4.8. The total front end bandwidth is not enhanced in any way since the TIA and the CTLE stage compensate each other, as illustrated in Figure 4.11, but this enables the use of a higher resistor value and therefore translates to lesser noise. In Equation 4.28, this is illustrated by the fact that f_{TIA} is reduced, thereby reducing the input referred noise. In this way, it appears that the transistor noise is somewhat more fundamental than the feedback resistor noise, even at low data rates.

Limits at higher data rates

At high data rates, the input referred noise contributed from the transistors is high enough that laser energy required to overcome it will be the dominant source of power consumption. In this case the optimal receiver will be optimized purely for noise and not it's own power

consumption, since it will be negligible. It can be easily derived from equation 4.29 that the optimal sizing for the input transistors will be $C_{in,TIA} = C_{PD}$. This yields the transistor noise limit, which, expressed in terms of photons per bit is:

$$n_{ph,min} = SNR\sqrt{32\pi\frac{C_{PD}}{6.4aF}\frac{f_{data}}{f_T}\frac{\gamma}{\alpha(1+\beta)}} \tag{4.31}$$

Limits in the low datarate case

At lower data rates, the energy will not necessarily be dominated by the laser. If we consider only the noise from the TIA transistors and the power consumption of the TIA and the laser, the energy per bit consumption of the link is:

$$E_{bit} = SNR\ V_{TX} I_{n,amp} T_{bit} + I_{TIA} V_{DD} T_{bit} \tag{4.32}$$

$$= SNR\ V_{TX} q\sqrt{8\pi\frac{(C_{PD}+C_{in,TIA})^2}{6.4aF \times C_{in,TIA}}\frac{f_{data}}{f_T}\frac{\gamma}{\alpha(1+\beta)}} + C_{in,TIA}(1+\beta)2\pi f_a V_{ov} V_{DD} T_{bit} \tag{4.33}$$

In this case, there is an optimal size for $C_{in,TIA}$. The lower the data rate, the smaller the input capacitance of the TIA will be in order to minimize power consumption for that stage. To obtain an analytic expression, we assume that $C_{in,TIA} << C_{PD}$, which leads to:

$$E_{bit,opt} = 3[\pi SNR\ V_{TX} C_{PD}]^{2/3}[V_{DD} V_{ov} \gamma k_B \Theta]^{1/3} \tag{4.34}$$

The surprising conclusion from Equation 4.34 is that the optimal energy per bit in this case does not depend on the datarate or the speed of the transistors f_T when the link energy is not dominated by the laser power.

E/b power laws

The limit between these two regimes is when we can no longer use the approximation $C_{in,TIA} << C_{PD}$ which is only valid when

$$4(\frac{SNR\ V_{TX}}{2V_{DD}V^*})^{2/3}(\frac{qV_{th}\gamma}{C_{PD}})^{1/3}\frac{1}{\alpha(1+\beta)} << \frac{f_T}{f_{data}} \tag{4.35}$$

With the photonics platform described in section 4.2, this leads to $\frac{f_T}{f_{data}} \sim 15$ which clearly states why 25Gb/s is in the laser limited regime, whereas 5Gb/s is in the full

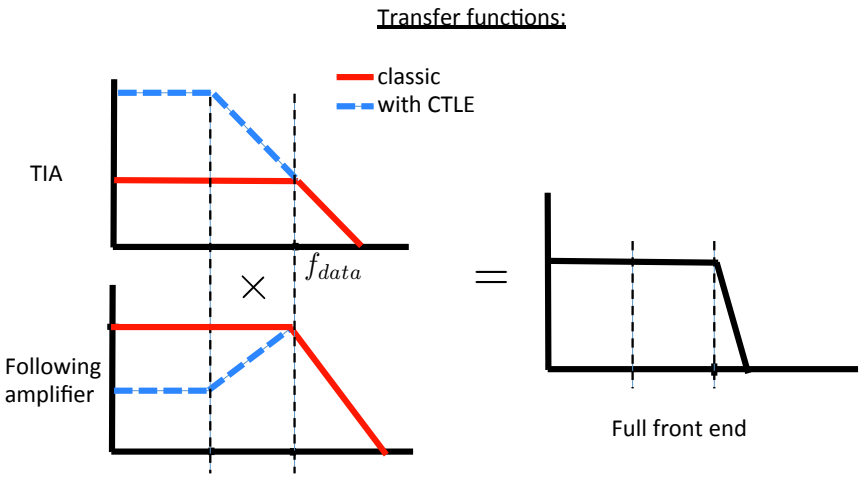

Figure 4.11: Ideal Transfer Function of A System with Equalization

link limited regime. The power laws for optimal Energy/bit of these different regimes is summarized in Table 4.3.

Table 4.3: Power laws for E/b limits dependence

Variable	TX dominated regime	TX and RX balanced regime
	Regimes defined by equation 4.35	
C_{PD}	1/2	2/3
V_{TX}	1	2/3
f_{data}	1/2	0
f_t	-1/2	0

We also express the different energies per bit in terms of the Landauer limit kT:

Table 4.4: Energy per bit in multiples of kT

Optics (low datarate)	$3[\pi \frac{C_{PD}}{6.4aF} \frac{V_{TX}}{V_{th}} SNR]^{2/3} [\frac{V_{DD}}{V_{th}} \frac{V_{ov}}{V_{th}} \gamma]^{1/3}$
Optics (high datarate)	$\frac{V_{TX}}{V_{th}} SNR \sqrt{32\pi \frac{C_{PD}}{6.4aF} \frac{f_{data}}{f_T} \frac{\gamma}{\alpha(1+\beta)}}$

As can be seen from table 4.4, the energy for communications can be expressed as a the Landauer limit kT times a number of inefficiencies expressed as ratios from ideal values for V_{TX}, C_{PD}, V_{DD}, V_{ov} and f_T

Figure 4.12 illustrates the two regimes of operation clearly.

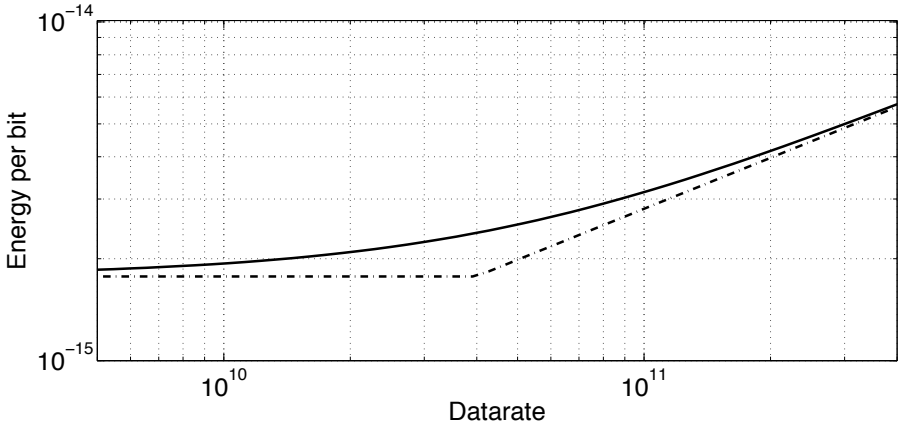

Figure 4.12: Energy per bit versus datarate, and the asymptotic curves from equations 4.31 and 4.34 ($V_{TX} = 40V;\ V_{DD} = 0.5V;\ V_{ov} = 50mV;\ SNR = 7; C_{PD} = 1$ fF; $f_T = 400$ GHz)

4.5 Observations in Scaling and Technology

With performance limitations arising from both the quality of the CMOS and photonic devices, this section aims to study the effects of an improved design platform with respect to optimized energy per bit. Following the previous analytic analysis, here we utilize the model and optimization procedure described in section 4.1, and apply it to different hypothetical technology platforms. This enables the capture of additional effects such as sampler energy not described in section 4.4. In doing so, we hope to target key bottlenecks in performance and potential for improvements in the next-generation of integration technologies.

Improvements in Photonics and Interconnects

Parasitics such as coupler losses and photodiode capacitance dominate the platform described in table 4.1 and limit the achievable energy efficiency. To study the importance of the photonic performance, we replace the existing metrics for coupler losses and photodiode capacitance, C_{PD} from 6.5dB/coupler and 20fF to 1dB/coupler and 3fF, respectively, implying $V_{TX} = 15V$. In addition, modulator efficiency as low as 1 fJ/bit have been demonstrated, justifying their omission from this analysis [18].

The results of the analysis are shown in Figure 4.13. As compared with the existing heterogeneous integration platform, using better photonics shows more than an order-of-magnitude improvement in link efficiency. Because the price to convert from the photonic to electrical domain, V_{TX}, is so cheap now, the optimized links at the various data rates are more receiver-performance limited, as expected intuitively.

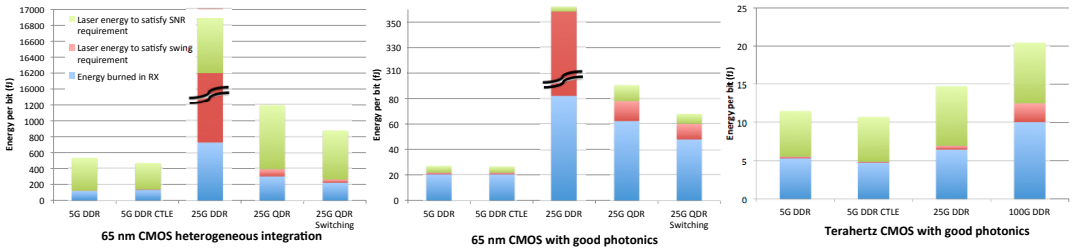

Figure 4.13: Technology Dependent Performance Prediction

Improvements in Photonics+CMOS

To push the boundary of integration technologies altogether, the case where the photonics and CMOS are both pushed to their bounds is studied. In particular, the same best photonic specifications from before is used, but now the technology node is scaled to reflect a theoretical f_T of 1THz. The results of the study are shown in Figure 4.13. For lower data rates, the performance improvement from scaling f_T from 150GHz to 1THz is observable but not drastic and stems mostly from the lower energy consumption of the samplers themselves and not the front end amplifier or the laser, as expected from the limits of section 4.4. For the 25G DDR case, however, the improvement is almost an order of magnitude since the faster amplifiers can provide gain at these speeds. Notice that the last column in this bar plot shows a *100G DDR* receiver, with a theoretical best end-to-end link efficiency of 20fJ/bit.

While the previous sections show the performance for given technologies, we can reverse the exercise to deduce the necessary technology properties for a given link efficiency. To achieve sub 1fJ/bit efficiency at 5Gbps and f_t=1000 Thz, this would require C_{PD}=200 aF, V_{DD}=0.5 V, V_{ov}=0.1 V and V_{TX}=10 V. These small photodiode capacitances would require such a small device that some sort of absorption enhancement would be necessary, such as a cavity or a metaloptic focusing scheme. At this point the link energy itself is so small that effort must be redirected to the energy overhead of peripheral blocks such as clock networks and bias generators.

The performance results for these higher data rates have another interesting trend – as the CMOS platform performance improves, the energy consumption of the receiver is mostly limited by the sampler itself. Because we have assumed a StrongArm topology for the sampler for all data rates of operation, the minimum achievable E/b of this sampler is far greater than the rest of the link put together. This yields the conclusion that within the confines of a better platform where photon efficiency is so high, using a simple gain stage such as an inverter as the sampler is more optimal than having a StrongArm or CML latch.

4.6 Ultimate limits

The photodetector capacitance, as seen from equations 4.34, 4.31 is one of the crucial lever that can be used to diminish the energy used by optical data links. Indeed, the lower the capacitance of the photodiode, the higher the voltage created by the absorbed photons will be, easing the SNR requirements on the photoreceiver end. The photodetector capacitance must also include the wire capacitance leading to the first amplification stage. For standalone photodiodes and long wirebonds, this means large capacitances on the order of $\sim 100fF$, which leads to high energy per bit. With wires having a capacitance of 200 $aF/\mu m$, getting rid of this extra capacitance has been the motivation for higher and higher integration of photonics and electronics, culminating in Silicon Photonics technology, where the photonics and the electronics sit on the same chip, allowing for very low interconnect parasitic capacitance. Heterogeneous integration has enabled photodetector capacitances as low as a few tens of fF [15], while homogeneous integration strategies have shown capacitances of just a few fF by putting the first transistors as close to the photodiode as possible and therefore almost entirely getting rid of the wire capacitance.

Optimal photodiode capacitance

All the noise analysis derivations that were performed in chapter 3 were done assuming a small signal regime for the receiver: $v_s << V_{ov}$. If this assumption breaks down, the transistor noise will no longer be dominated by the bias current but by the signal current, and will therefore be imposing a constraint on the photon current identical to the photon shot noise limit. Given the fact that at least $2SNR^2$ photons must be used to satisfy the quantum shot noise condition, the signal voltage in that limit is $v_s = 2SNR^2q/(2\pi C_{in})$ leading to the small signal condition $\frac{2SNR^2q}{2\pi V_{ov}} << C_{in}$. As shown previously $C_{in} < 2C_{PD}$. This leads to the conclusion that lowering the capacitance of the photodiode will only be useful down to a certain capacitance:

$$C_{PD,min} = 6.4aF\frac{V_{th}}{V_{ov}}\frac{SNR^2}{2\pi} \tag{4.36}$$

For a bipolar transistor where $V_{ov} = V_{th}$, and for a SNR of 7 (BER $\sim 10^{-12}$) this leads to $C_{PD,min} \sim 50\ aF$. For a MOS transistor with an overdrive voltage of 0.1 V, this means $C_{PD,min} \sim 13\ aF$. While making photoreceivers with such small capacitances and high responsivity is hard, it is not impossible with appropriate light management.

For example, in Germanium, once photons above the direct bandgap are used, the absorption length lies roughly at $\sim 1\mu m$. This means that if photodetectors with signficantly smaller dimensions are to be used, they will have to be accompagnied by an absorption enhancement scheme in order to maintain acceptable quantum efficiency such as a dielectric cavity [19] or a metaloptics focusing device.

Absorbing light in small values with a dielectric cavity

Famously, light in a dielectric medium cannot be focused beyond the diffraction limit, meaning that the minimal possible mode volume of a cavity will be $(\frac{\lambda}{2n})^3$, (where λ is the size of the wavelength in vaccuum and n is the refractive index of the material). If a small absorbing volume of material V_{abs} is placed in a half wavelength cavity, and neglecting all other loss mechanisms, the quality factor Q of the cavity that will allow for full absorption of the light is

$$Q = Q_{int,mat}\frac{(\frac{\lambda}{2n})^3}{V_{abs}} \tag{4.37}$$

where

$$Q_{int,mat} = 2\pi\frac{n}{\alpha\lambda} = \frac{\epsilon'}{\epsilon''} \tag{4.38}$$

is the intrinsic Q factor of the material, n, α, ϵ', ϵ'' are respectively the refractive index, the absorption coefficient, the real and imaginary part of the permittivity of the absorbing material. For Germanium, $Q_{int} \sim 16$, and most semiconductors will take roughly this value above their direct bandgap.

We can therefore plot out the required quality factor of a $(\frac{\lambda}{2n})^3$ cavity for full absorption of incoming light with respect to the dimension of a cube of absorbing material enclosed in the cavity.

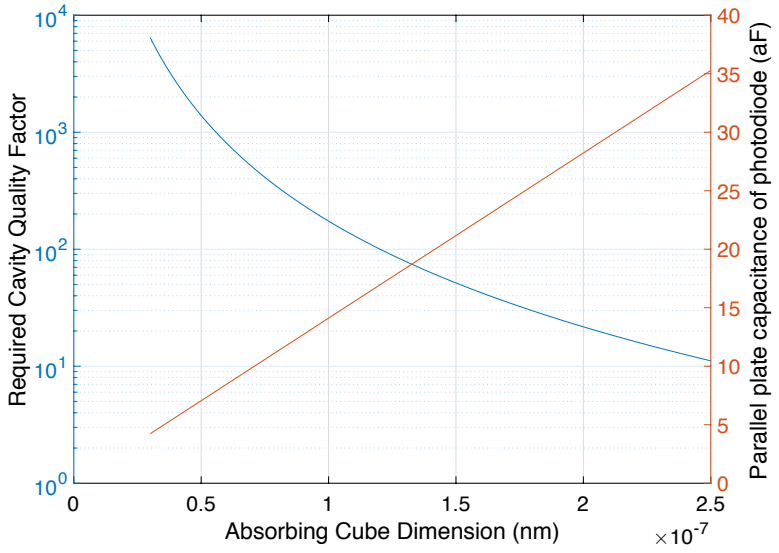

Figure 4.14: Cavity quality factor required for efficient light absorption by a small volume of Germanium ($\lambda = 1500nm$) and parallel plate approximation of the Germanium's capacitance

Reaching the optimal PD capacitance

As pointed out earlier, the optimal photodiode capacitance lies in the few 10's of attofarads in order to allow for the incoming photons to induce a high voltage swing at the input of the first transistors. If we assume a very simple parallel plate capacitor model, a cube of Germanium $100nm$ in size will have a capacitance of $\sim 14aF$ and will require a cavity with a Q-factor ~ 200 according to figure 4.14. Naturally the parallel plate approximation is underestimating the actual capacitance of the photodiode, and is not taking into account the capacitance of the required wires leading to the first transistor, nonetheless this is clearly in the realm of the possible: optical cavities with much higher Q factors have already been demonstrated, and semiconductor devices are made at much smaller dimensions than required here.

If even smaller absorbing volumes are to be used, it may be impossible to neglect all other losses mechanisms beyond the absorption from the photodiode itself, such as optical losses in the contacts leading to the device. Another easier limit of optical cavities to quantify is the limit imposed by the duration of energy storage. Indeed an optical cavity used for communications should not store energy longer than the duration of a bit in order to avoid inter-symbol interference. At 100 Gbps, this limits the cavity Q to $\sim 10^3$ if 1550 nm light is used. Beyond that speed a scheme capable of concentrating light beyond the diffraction must be used, such as metaloptics.

The path towards the quantum limit

We summarize the different limits in figure 4.15, depending on the photodetector capacitance (and categorized on the level of integration used), and the energy cost of the photons. Modern optimal communications systems today lie in the top right of the graph in the picoJoule range, due to the large losses in the optical path, as well as very high capacitance of the photodiodes due to the lack of extremely tight integration.

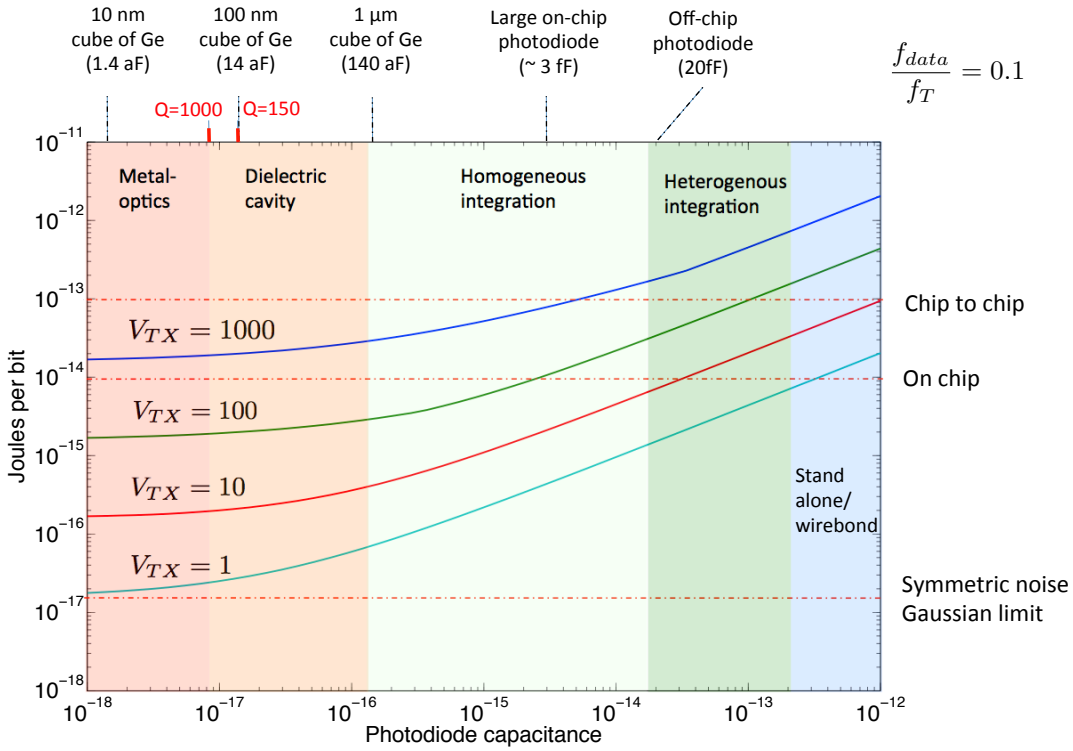

Figure 4.15: Energy per bit versus photodiode capacitance, for different wall plug efficiencies of photons at the photoreceiver (V_{TX} defined in equation 4.26, $V_{ov} = 0.1V$, $V_{DD} = 0.3V$). In dotted red lines the energy objectives necessary for chip to chip or on-chip optical interconnects to be viable (from [5])

This shows the great potential for improvement in optical interconnect energy usage, provided smaller photodiodes and higher integration are used in order to strongly diminish the capacitance lying before the first amplifying transistor.

Chapter 5

Phototransistors

Phototransistors are semiconductor devices that take an optical input, and convert that signal to an electrical signal in the same way as a photodiode while also amplifying it using transistor action. In that sense they are a monolithically integrated version of a photodiode and transistor. There a large variety of different phototransistor flavors, which closely follow the different types of transistors: photo-JFETs [20, 21], photo-MOSFETS [22, 23], bipolar phototransisors (BPTs) [24, 25].

As seen in chapter 3, the capacitance of the photodiode is crucial in determining the sensitivity of the front end. In this respect, integration of the first stage of gain and the photodiode makes a lot of sense. Indeed, if one can get rid of all the wire parasitic capacitance, one can hope to increase the sensitivity of the front end, and reduce the power consumption of the link. Nonetheless this reasoning has several pitfalls that will be covered in detail in this chapter. It will be shown that the biasing of phototransistors is problematic, that phototransistors require more stringent equalization schemes to overcome kTC noise than TIA front ends as pointed out in chapter 3, and finally that there is a fundamental mismatch in size requirements for the two functions of the phototransistor: light absorption and current amplification. While ways of dealing with the last problem will be presented, it is still hard to justify how a phototransistor may be useful for Telecom applications. Part of the work presented here has been published in [26, 27].

5.1 How do phototransistors work?

From a first order perspective, the operation of phototransistors is very similar to that of photoconductors with gain: photons are absorbed and create electron and hole pairs in the semiconductor, as illustrated in fig 5.1. The holes (assuming the electron is the gain carrier) then stay in place while electrons can go from the cathode to the anode several times. Only one conduction electron per hole can traverse the device at a time. The current coming out of the device is therefore

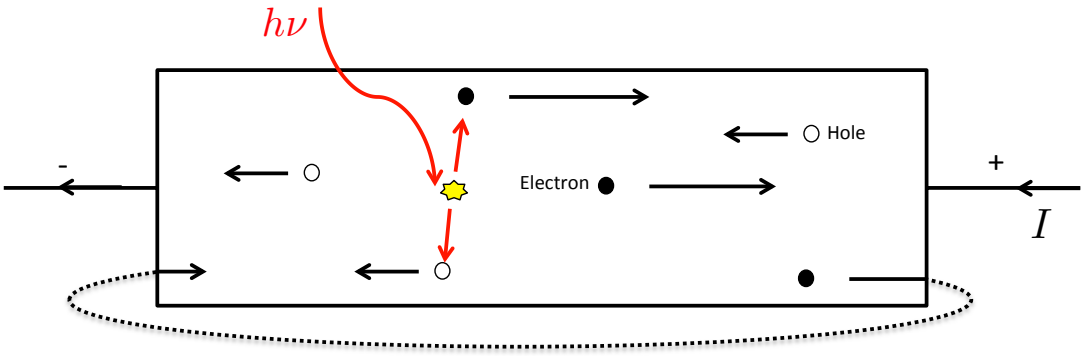

Figure 5.1: A simple photoconductor schematic. Holes and electrons created by the incoming light provide current carriers that will drift with the applied bias. Electrons (and holes) may circulate through the device several times

$$i_{out} = i_{ph} \frac{\tau_{transit}}{\tau_{lifetime}} \tag{5.1}$$

where $\tau_{transit}$ is the transit time of the electrons through the photoconductor and $\tau_{lifetime}$ is the lifetime of the holes in the device and i_{ph} is the photon current. For the different phototransistors mentioned previously, the transit time would be the time for the conduction electrons to go through the channel in the MOSFET and JFET case and the time to go from the emitter to the collector for the BPTs.

Naturally this is a very simplified view of photoconductors: holes also contribute to the current in the device, but due to their lower mobility their contribution will often be less than that of the electrons, and it is usual to approximate the electrons as the carriers providing most of the current.

The basic operation of these different devices is described here.

Photo-MOSFET

The operation principle of a photo-MOSFETs is as such: the gate of the device is made of a semiconductor with a bandgap energy bellow that of the photons used to communicate the data, such at Germanium for telecom wavelengths. The body of the transistor itself is made of a larger bandgap material, so that it is transparent (for example Silicon). The semiconductor in the gate must be doped to be of an opposite type of the body, such as depicted in figure 5.2, and the device is biased such that the channel of the transistor is slightly inverted. The band bending created by the biasing of the gate creates a junction

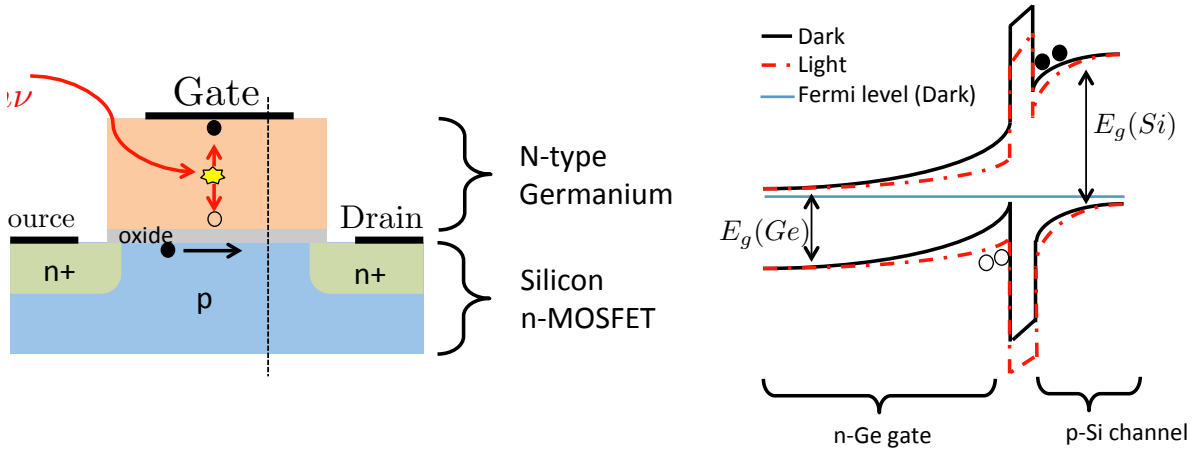

Figure 5.2: Photo-MOSFET schematic and band diagram of a cross section of the device with and without illumination

that will separate photocarriers created by the incoming light. Holes (in the case of an n-doped gate) are then trapped at the gate oxide and invert the channel further, increasing the source drain current. Once again, each hole trapped at the oxide interface can support a single conduction electron at a time. The device turns off as holes recombine in the germanium.

Photo-JFET

For the photo-JFET in [21], the gate of the device is made of Germanium grown directly on Silicon and its operation relies on discontinuities in the valence band between Germaniun and Silicon. When light is absorbed in the germanium, the electrons can diffuse to the gate and be swept by the drain, but the holes are stuck in the Ge by the valence band discontinuity. These holes provide a bias that modulates the depletion region and therefore the channel of the JFET , as shown in figure 5.3. Once again, each hole trapped at the semiconductor interface can support a single conduction electron at a time. The device turns off as the holes either recombine or jump over the band discontinuity.

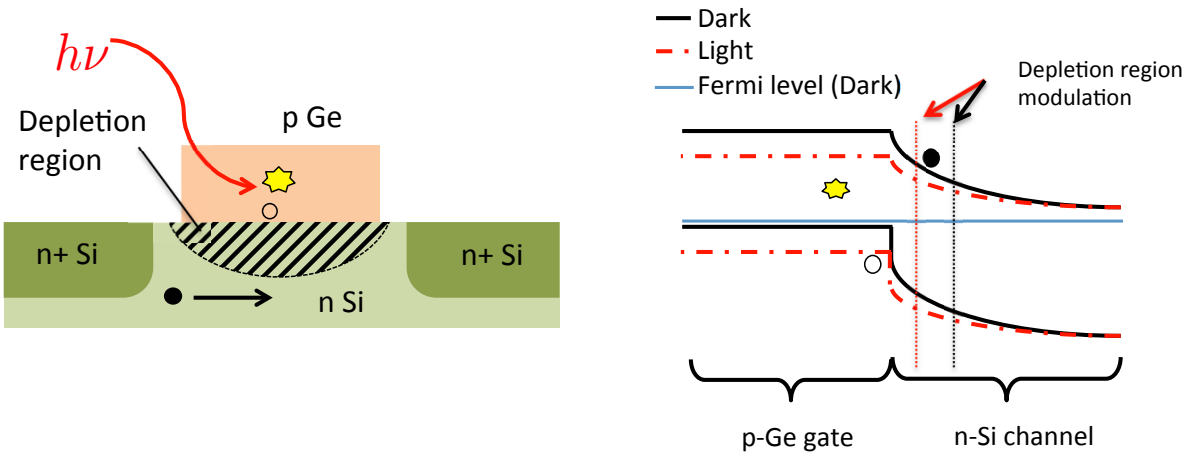

Figure 5.3: Photo-JFET schematic and band diagram of a cross section of the device with and without illumination

Bipolar-phototransistors

Bipolar phototransistors will be described in much further detail in 5.2. In a BPT, the entire device is made of a photosensitive semiconductor. Light is absorbed in the base collector region, and the holes are swept to the base, where they are trapped. They also provide a positive bias to the base, which lowers the barrier for electrons to flow from the emitter to the collector, creating the amplified current. Once again, only one electron can be traversing the device at any time for each hole trapped in the base. The device turns off as the holes either recombine with electrons or overcome the valence band barrier and reach the emitter.

Since most of these different phototransistors actually work in very similar ways, only the bipolar photo-transistor will be studied in detail. The results and conclusions can be easily generalized to all other type of phototransistors though.

5.2 Bipolar photo-transistors (BPTs)

The idea of using bipolar junction transistors (BJTs) as photodetectors with gain was introduced by Shockley at the same time he created his first purely electrical device [28]. Bipolar phototransistors (BPTs) have since then been considered countless times [29, 30,

31, 24, 32] as alternatives to p-i-n and avalanche photodetectors, yet have failed to actually replace them for telecom applications.

The bipolar transistor itself has seen many improvements since Shockley's first iteration, one of the most noticeable one being the introduction of a hetero-junction as first suggested by Kroemer [33]. Vertical scaling and smart device design to reduce stray capacitance are also responsible for the continuously improving performance of Heterojunction bipolar transistors (HBTs). They have been demonstrated in the Silicon-Germanium and III-V material system with speeds up to 710 GHz [34], have been theorized for speeds up to the Terahertz range [35, 36] and are currently used in many RF applications as a crucial element in BiCMOS technology [37].

Basic bipolar transistor theory

An electrical BJT is a semiconductor device composed of three doped region: the emitter, the base and the collector. The base is doped to be of the opposite type than the emitter and the collector, so that the transistor can be n-p-n if the the base is p-type, or p-n-p if the base is n-type. Usually, the n-p-n configuration has the highest performance, as electrons have higher mobility than holes. Here, unless stated otherwise, it is assumed to be n-p-n.

The most common biasing scheme is the common emitter scheme, where the emitter is grounded, and the collector is biased to a positive voltage. The resultant band diagram is depicted in figure 5.4. The gain β happens between the input base current and the output collector current:

$$I_c = \beta I_b \tag{5.2}$$

DC gain

In short base BJTs, the base current is composed mostly of holes that diffuse from the base into the emitter and recombine at the emitter contact. The collector current, on the other hand, comes from electrons that have diffused from the emitter all the way through the base and are collected by the strong field at the base-collector junction. Therefore the gain can also be defined as the ratio of electrons to holes flowing between the base and the emitter, and is thus strongly defined by the properties of that junction. In homojunction BJTs, the gain is defined by the ratio of doping levels in the base N_{ab} and emitter N_{de}, the ratio of the lengths of the base W_B and the emitter W_E, and finally the ratio of the diffusion coefficients for holes in the emitter D_{he} and electrons in the base D_{eb}. In heterojunction transistors, the emitter is made of a semiconductor with a higher bandgap, so that the energy barriers for the holes to diffuse into the emitter is greater, which adds another term

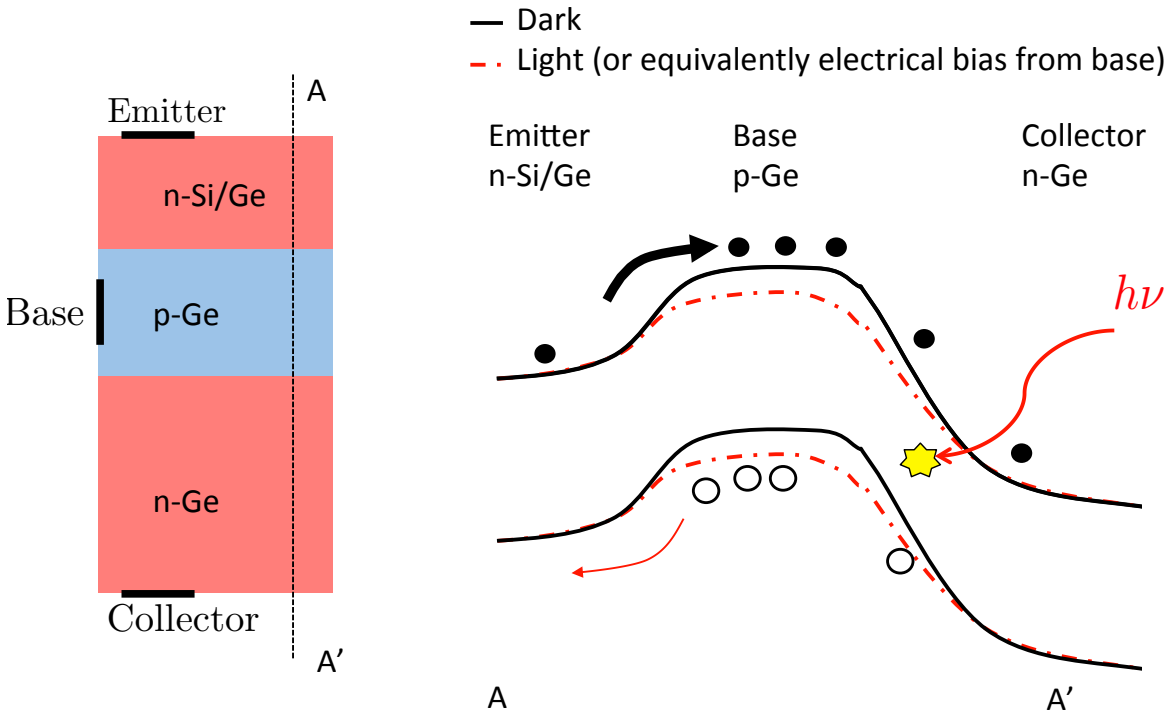

Figure 5.4: HBT/BJT schematic and band diagram of a cross section of the device with and without a base bias. This bias can be electrical as in electrical BJTs, or optical, as for a phototransistor

to the gain [38].

$$\beta_{homo} = \frac{W_E N_{de} D_{nb}}{W_B N_{ab} D_{pe}} \tag{5.3}$$

$$\beta_{hetero} = \frac{W_E N_{de} D_{nb}}{W_B N_{ab} D_{pe}} \exp \frac{\Delta E_G}{kT} \tag{5.4}$$

where ΔE_G is the difference in bandgap between the emitter and base. This means that in order to have high gain, a BJT must have high emitter doping, a relatively thin base and that a higher bandgap emitter is advisable. Nevertheless, these requirements are far from sufficient to ensure that the device can operate well at high speeds.

High frequency operation

At high frequencies, the speed of the device will depend on the different capacitances that need to charge and discharge. There are two main capacitances that must be taken into account: the static capacitances and the diffusion capacitance. The former is composed

of the capacitances created at the emitter-base and collector-base junctions, and can be easily calculated with the classical parallel plate capacitance formula $C = \frac{\epsilon A}{d}$. The diffusion capacitance is a more subtle charge storage mechanism, and comes from the fact when electrons diffuse as minority carriers from the emitter to the collector, they carry a certain amount of charge which must be compensated by extra holes in the base. This leads to an extra term in the base-emitter capacitance which depends on the amount of current flowing through the base. The charge stored can be written as

$$Q_{stored} = I_C \tau_F \tag{5.5}$$

where τ_F is time it takes an electron to transit from the emitter to the collector. Since I_C results from the electron current being injected from the emitter in a manner analogous to a p-n junction, we have $I_C = I_0 \exp \frac{qV_{BE}}{kT}$, and it can be therefore easily derived that

$$C_{diff} = \frac{\partial Q}{\partial V_{BE}} = \frac{I_C \tau_F}{V_{th}} = 6.4aF \frac{\tau_F I_C}{q} \tag{5.6}$$

The main effect of the the capacitances is to reduce the gain at high frequencies. This is best described by the the value of the transition frequency f_T which the value at which the current gain drops to unity, and can be calculated as [38]:

$$f_T = \frac{1}{2\pi r_\pi (C_{static} + C_{diff})} = \frac{1}{2\pi(\tau_F + C_{static}\frac{I_C}{V_{th}})} \tag{5.7}$$

where r_π is the small signal resistance of the base-emitter pn junction.

It is clear from 5.7, the transit time τ_F is a hard limit to the speed a bipolar transistor can work at and provide gain. The transit time can be expressed as the sum of the individual time through the different sections of the transistor. The most important contributors are the base transit time τ_B and the base-collector depletion region transit time τ_{BCD}. To first order, these can be expressed as:

$$\tau_F \sim \tau_B + \tau_{CBD} \tag{5.8}$$

$$\tau_B = \frac{W_B^2}{2D_{nb}} \tag{5.9}$$

$$\tau_{CBD} = \frac{W_{BCD}}{2v_{sat}} \tag{5.10}$$

The transport time across the base results from diffusion and is as expected dictated by the diffusion constant of holes, and the length of the base. The transit time through

the base-collector depletion region, on the other hand results from the drift of the electrons thought the high electric field in the junction. If the junction is biased strongly enough (which is usually the case), the electrons drift at the saturation velocity v_{sat}. A careful analysis shows that the charge compensation only needs to happen in half of the junction, therefore explaining the factor of 2 in the denominator of equation5.10 [39].

Circuit model and noise sources

The simplest model that encompasses the most important considerations for speed and noise calculation in a small signal framework is the hybrid-π model, shown on figure 5.5 (the figures also includes the current sources that illumination would add, and which would obviously not be present for a purely electrical BJT). C_π represents the capacitances between the emitter and includes the base emitter static junction capacitance $C_{J,BE}$ as well as the diffusion capacitance C_{diff} from 5.6. C_μ is the base/collector junction capacitance $C_{J,BC}$, r_π is the base emitter dynamic resistance, g_m is the transconductance of the transistor. All can be expressed as:

$$r_\pi = \frac{\partial V_{BE}}{\partial I_B} = \frac{kT}{qI_B} = \frac{\beta}{g_m} \tag{5.11}$$

$$C_\pi = C_{diff} + C_{J,BE} = \frac{I_C \tau_F}{V_{th}} \tag{5.12}$$

$$C_\mu = C_{J,BC} \tag{5.13}$$

$$g_m = \frac{\partial I_C}{\partial V_{BE}} = \frac{qI_C}{kT} \tag{5.14}$$

The main sources of noise come from the base current shot noise and the collector current shot noise. Their noise power follow the form given in chapter 2:

$$I_{n,B}^2 = 2qI_B \tag{5.15}$$

$$I_{n,C}^2 = 2qI_C \tag{5.16}$$

Bipolar phototransistors

Bipolar phototransistors work in a very similar fashion than electrical bipolar transistors. The only difference is that the absorbed light creates an additional current source, which derives not from an electrical input but an optical one. The photons absorbed in the base-collector depletion region create electron and hole pairs which are separated by the strong field. The holes accumulate in the base and create the voltage bias that increases the collector current. Photons absorbed in other parts of the device may contribute to base current if the holes diffuse their way to the base, but that process can be much slower since it is driven by a diffusion process. For data communication, the rate at which the photons must be collected must at least be as fast as the datarate itself. Additionally the

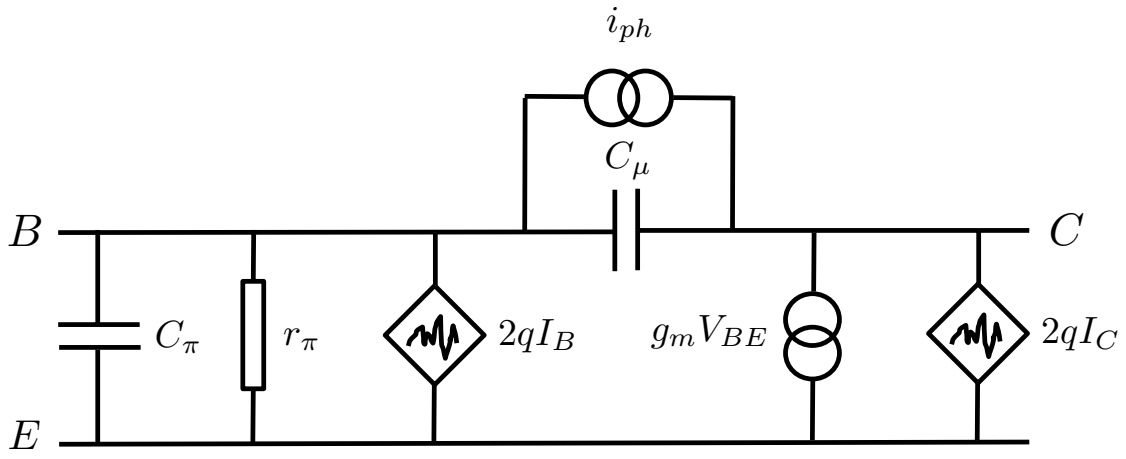

Figure 5.5: Hybrid-pi circuit model of a bipolar photo-transistor, including base and collector current noise sources.

holes can also recombine before reaching the base and be lost. This is why it is strongly desirable to capture photons in the base collector depletion region.

Noise analysis

The noise analysis was performed in chapter 3, and it is shown that the sensitivity of a bipolar phototransistor is (for a low impedance load where the voltage of the collector stays constant):

$$n_{ph} \sim \frac{2SNR}{\eta}\sqrt{n_{ph}I_2 + 2\pi\frac{B_{BPT}}{f_{data}}\frac{C}{6.4aF}I_2 + 2\pi\frac{C}{C_{diff}}\frac{C}{6.4aF}\frac{f_{data}}{f_{t,m}}I_3} \tag{5.17}$$

where B_{BPT} is the bandwidth of the phototransistor and:

$$C = C_B + C_{diff} = C_{J,BE} + C_{J,BC} + \frac{\tau_F I_C}{V_{th}} \tag{5.18}$$

$$B_{BPT} = \frac{1}{2\pi C r_\pi} \tag{5.19}$$

$$f_{t,m} = \frac{1}{2\pi\tau} \tag{5.20}$$

The first term in 5.17 is due to the photon shot noise, the second comes from base current shot noise, while the third is the resultant of collector current shot noise. As described in section 4.3, it is possible to mitigate the effect of the base shot noise by

decreasing the bandwidth of the phototransistor and compensating for the lower bandwidth by using an equalization stage after the phototransistor. If we assume that this technique is used and that the photon shot noise is negligible (which is almost always the case), we are left with only the collector shot noise, which is optimized for $C_{diff} = C_B$:

$$n_{ph,opt} \sim \frac{2SNR}{\eta}\sqrt{8\pi\frac{C_B}{6.4aF}\frac{f_{data}}{f_{t,m}}I_3} \tag{5.21}$$

5.3 Decoupling gain and absorption: a new type of BPT

The issue with classical BPTs

In classical bipolar phototransistors, the fact that light is absorbed in the base-collector depletion region comes with certain constraints. Indeed, the transfer time τ_F of electrons through the device, which ultimately limits it's speed, includes the base/collector transit time $\tau_{BC} = \frac{W_{BC}}{2v_{sat}}$, (where W_{BC} is the physical length of the base collector junction) as described in 5.9. What this implies is that there is an inherent between high speed and long absorption region. The absorption length of telecom wavelengths in most materials is on the order of several microns, whereas the base/collector junction in modern transistors is on the order of tens of nanometers. There is therefore a mismatch of several orders of magnitude between the two, which implies that in a conventional topside illumination structure where the absorption length and the base collector junction are on the same axis it is impossible to have a phototransistor that is both fast and has high absorption efficiency.

This problem also appears for classical pin photodiodes, and has been alleviated by moving the absorption length to a different dimension perpendicular to the junction [40], which effectively decouples the absorption length from the junction length. This trick can also be employed for a BPT, but it implies a large area for the base/collector junction, which in turn means a high base capacitance. Indeed the base/collector junction has a very large capacitance per unit area since it must be short to ensure high speeds. This has adverse effects for the device sensitivity, as seen in equation 5.21.

In short, for a conventional BPT structure, the absorption volume cannot be made large enough without sacrificing speed and sensitivity.

Decoupling gain and absorption

Selectively implanted collectors (SIC) in electrical HBTs, and using it for phototransistors

The layout of BJTs and HBTs has considerably evolved since Shockley's implementation, and modern HBTs used in BiCMOS look nothing like his first device. One crucial improvement has been the implementation of a selectively implanted collector (SIC) [41] which keeps the area of the base/collector junction comparable to the emitter/base junction and therefore minimizes it's capacitance, while still allowing the base to have a large area so that it can be electrically connected as shown in figure 5.6. The SIC is done by implanting through the base before the emitter is epitaxialy grown. On the sides of the SIC, a junction between the base and the subcollector is present with a depletion region much longer than the base/SIC junction, and therefore much lower capacitance.

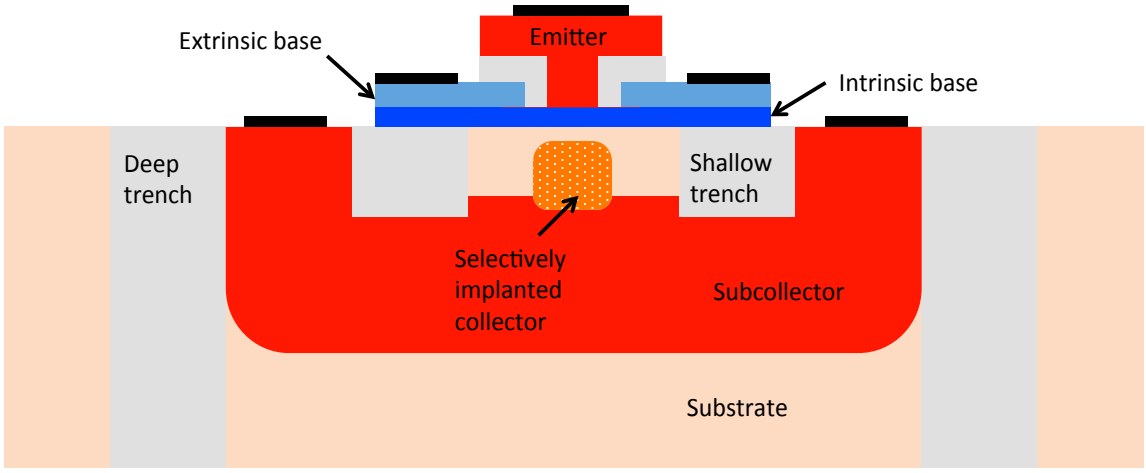

Figure 5.6: Schematic of modern high end heterojunction bipolar transistor. The selectively implanted collector enables short transit time and low capacitance.

Using this junction as the absorption region is a golden opportunity for BPTs: indeed it has a large absorption volume, yet its capacitance is very low. Any photons absorbed in this region will directly behave as base current, just as if they had been absorbed in the base/SIC junction. Holes will drift to the extrinsic base, where they will become majority carriers and bias the base positively. Naturally the photocarrier collection is slower than if it were happening in the base/SIC region since they have to transit through a longer junction, but the speed of the transistor itself which is defined by the length of the junction under the emitter and is the true bottleneck, is virtually unchanged. The only requirement for photocarrier collection is that it be relatively faster than the bit duration.

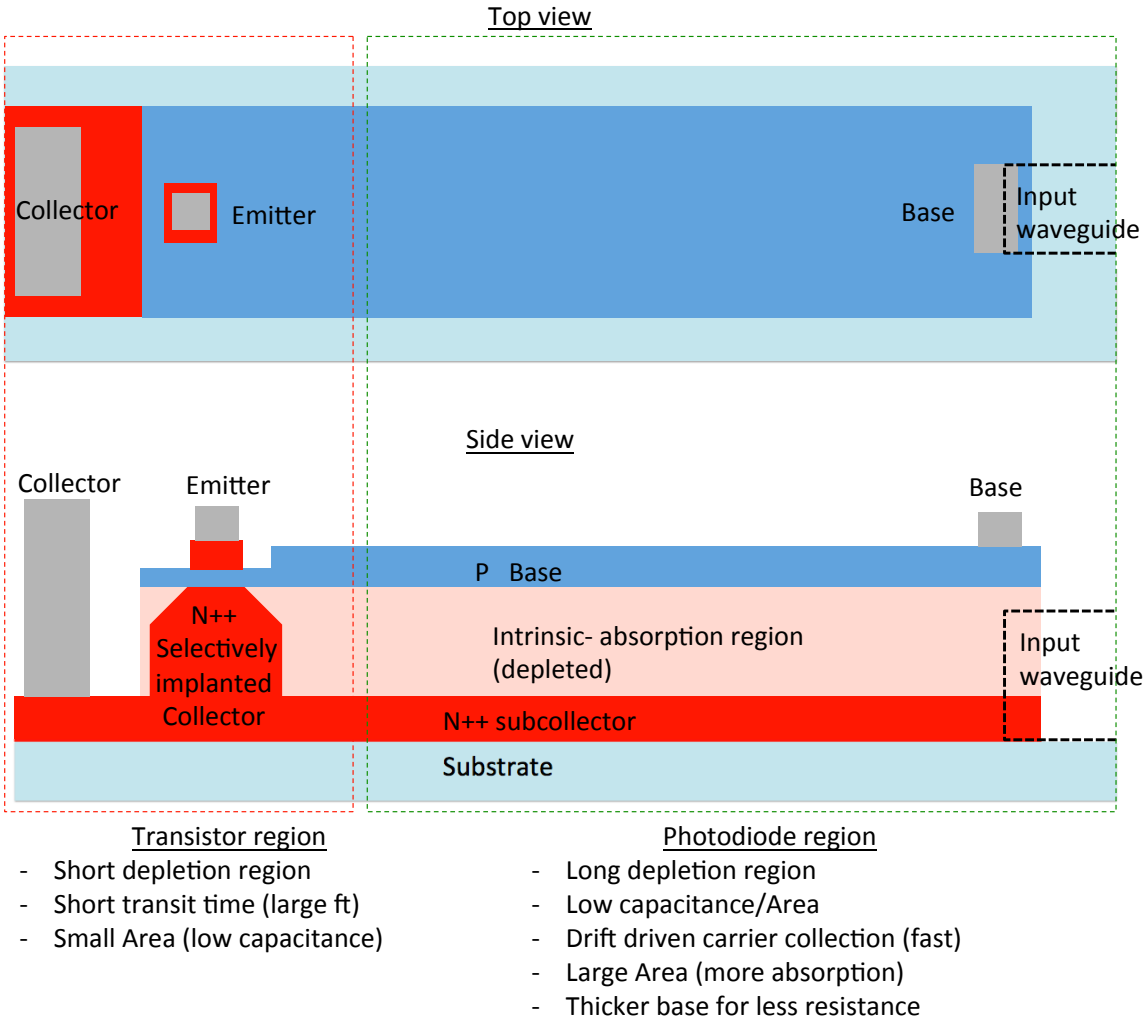

Figure 5.7: Schematic of an optimized BPT with decoupled absorption and gain region

This effectively decouples the gain region of the BPT from the absorption region, enabling the detector to have low capacitance and maintain high speed while keeping a large absorption volume, as depicted in figure 5.7. The area of the base/subcollector junction can be made much larger than the base/SIC junction before the extra capacitance starts strongly affecting the device performance.

For example, a typical capacitance for a modern 300 GHz HBT in 130 nm technology (emitter width) [37] will have a total base depletion capacitance of roughly 38 fF/μm^2, whereas a junction 300nm deep in Germanium has a capacitance of only 0.47 fF/μm^2. The depth of the subcollector/base junction that serves as the absorption region could also be made deeper to further reduce it's capacitance, the limit being the speed at which photocarriers have to be collected, and the length a SIC can practically be made. Naturally

if one wishes to work at the telecom wavelength, a semiconductor (such as Germanium) that absorbs at the appropriate wavelength must be used to fabricate the device.

Using the subcollector/base junction as the photon absorption region opens the door to a fast, low capacitance phototransistor with a large absorption volume. While a material system change is necessary in order to work with telecom wavelengths, the basic technology brick is already present with modern HBTs, offering its self-aligned fabrication methods and device understanding.

Proof of concept simulation results

In order to illustrate and demonstrate the operating principle of the device, it was simulated using a drift diffusion solver [42]. To demonstrate the effect of the SIC, a BPT was simulated with and without the SIC implant and the speed response of the device to an optical excitaiton was calculated. The device simulated was made of Germanium grown on a n-doped (1e19/cm3) Silicon substrate serving as the sub-collector. The germanium layer simulated was 400nm thick, and undoped, except for the top 50nm, which served as the doped base layer (p-type 2e18/cm3). The emitter window was simulated to be 100nm made. The SIC implant is performed through the emitter window before the deposition of the emitter, which is made of highly n-doped poly-Silicon (2e19/cm3). The lateral size was 2 μm. The collector was positively biased to 0.3V an a base current was applied to reach the optimal bias point. For the purpose of this proof of concept the capacitances of the contacts were assumed to be zero, which is unrealistic and will be discussed in 5.4. The frequency response to an optical signal was calculated for both a device with and without the SIC and the results are shown in figure 5.8.

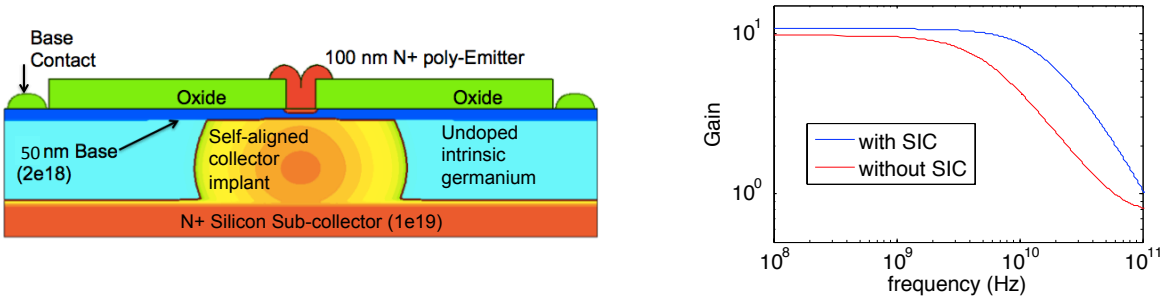

Figure 5.8: Comparison of speed response with and without the selectively implanted collector

It is clear that the gain-bandwidth of the device is drastically improved with the SIC. Nevertheless the absoption volume is roughly the same for both devices. This proves that SIC can be used to drastically improve the capacitance and speed of BPTS.

In order to further demonstrate the breakdown of the gain-bandwidth and optical absorption tradeoff, a similar device was simulated where the lateral size of the device was modified and the transition frequency f_T was calculated. The results are shown in figure 5.9, and clearly show that the absorption length of the device can be made longer with only a minor effect on the device f_T.

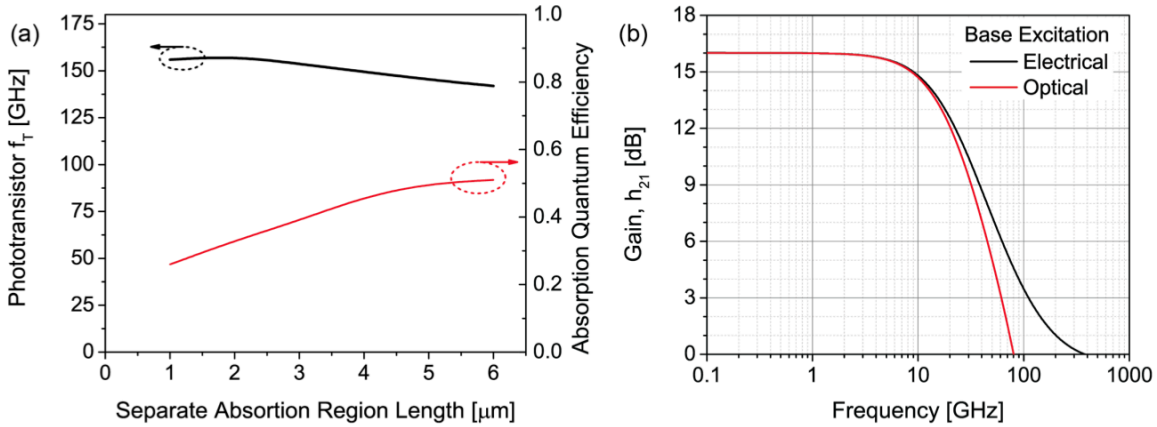

Figure 5.9: (a) Electrical transistion frequency and optical absorption efficiency for BPTs with different absorption lengths, (b) Gain versus frequency for a 1μm long device for an optical and an electrical excitation

5.4 Remaining issues with phototransisors

While the optimized BPT presented in 5.3 solves one of the major limitations of classical phototransistor by decoupling the gain and the absorption region, a number of issues still plague phototransistors if they are to replace photodiodes for telecom applications.

First comes the inherent difficulty of making a good transistor with a low bandgap material such as Germanium, which is very susceptible to avalanche latchup, but these difficulties can be engineered out and are not fundamental in nature.

Other issues though are more systematic and can only be solved by radical changes in in the system architecture.

Biasing

The first issue concerns the biasing of the transistor: in order for the phototransistor to have an acceptable speed and therefore noise performance, the base must be biased with a certain amount of current to reach it's optimal operating point. This usually means having a current source connected to the base with a wire. The entire point of using a phototransistor was to get rid of any wire that contributes to capacitance on the photoreceiver, so bringing in a wire to provide the bias current mostly defeats the purpose of making a

phototransistor.

Another way to bias the phototransistor is to use light itself. Indeed since the device is photosensitive by design, one can imagine using a close by light source to bias the device to the optimal point. Naturally this adds significantly to the complexity of the system. Nevertheless it is expected that such a device would be used in an environment where transmitters and therefore light sources would also be present, so that taping into these sources might not be excessively complicated.

Self biasing from the DC component of the light present in the datastream is also an option, but poses serious challenges due to the fact that their could be an important shift in the bias point depending on the data: indeed long runs of ZEROs or ONEs of an unbalanced stream of data will shift the "instantaneous" DC point of the data stream. Having a controlled DC component to the stream of photons is also highly undesirable as the energy cost of generating those photons will be very important.

kTC noise

The sensitivity of multiple different front ends including the TIA and the BPT were derived in chapter 3. The results are reiterated here:

$$n_{ph,TIA} \sim \frac{2SNR}{\eta}\sqrt{n_{ph}I_2 + 4\pi\frac{C}{6.4aF}\frac{B_{TIA}}{f_{data}}\frac{1}{G_{int}}I_2 + 4\pi\frac{C}{6.4aF}\frac{C}{C_{ox}}\frac{f_{data}}{\tilde{f_T}}\gamma I_3} \tag{5.22}$$

$$n_{ph,BPT} \sim \frac{2SNR}{\eta}\sqrt{n_{ph}I_2 + 2\pi\frac{B_{BPT}}{f_{data}}\frac{C}{6.4aF}I_2 + 2\pi\frac{C}{C_{diff}}\frac{C}{6.4aF}\frac{f_{data}}{f_{t,m}}I_3} \tag{5.23}$$

The second term in both expression comes from the base current shot noise for the BPT and the feedback resistor Johnson noise for the TIA, and are often referred to as kTC noise. It's value is inversely proportional to the value of the emitter-base dynamic resistance of the BPT and the feedback resistor in the TIA. As illustrated in 4.3, they can be reduced by decreasing the bandwidth of the front end which can be done by increasing the value of the feedback resistor in the TIA case, and increasing β in the BPT case. Nevertheless it should be noted that in the case of the TIA, the negative feedback allows for significant bandwidth enhancement for the same resistance value, leading to a noise power lower by a factor $\frac{G_{int}}{2}$. This implies that in order to reach the same kTC noise performance, the bandwidth must be significantly lower for the BPT, and the DC gain equally higher. High DC gain can cause a problem of dynamic range if there is a strong DC imbalance of the signal, such as when a long stream of ONES or ZEROS in the data happens.

In short the kTC noise is much worse for a BPT than a TIA scheme. Naturally the TIA scheme requires more than just one device, and therefore will most likely have a slightly

higher capacitance. Nevertheless if these are very tightly integrated on the chip the wires can be small enough that they only contribute a minor part of the total input capacitance, and the overall performance of the TIA scheme will be superior.

5.5 Conclusion

The exploratory study of using phototransitors for telecom applications has yielded a new type of phototransistor where the absorption medium and the gain region are decoupled, enabling fast devices with low capacitance and a large absorption region. These new phototransistors could have a variety of applications outside of the communications world such as in imaging. Nevertheless, for interconnect applications, since the main purpose of using a phototransistor is to keep the capacitance of the device as low as possible by avoiding attaching any wires to the photosensitive part of the device, this unfortunately leads to many complications in the design of the system. First off a bias current is necessary for the phototransistor to be at the optimum operation point, and while a light biasing scheme is possible, it seems like a very impractical solution. Additionally this prohibits the use of feedback schemes which can greatly enhance the bandwidth and consequently reduce the kTC noise. Ultimately, a photodiode and a transistor are two devices with very different requirements from many different standpoints (bandgap, size) and while integrating them together does offer the possibility of getting rid of a wire capacitance, it also carries major system design issues. Overall, if the devices can be tightly integrated, the capacitances of short wires connecting them can be virtually negligible and it is very difficult to see a path where phototransistors would provide better performance than separate photodiodes and transistors.

Chapter 6

Shape optimization for Silicon Photonics

As demonstrated throughout this work, high efficiency optical communication require extremely tight integration of electronics and photonics in order to reach low capacitances. One of the most promising platforms for this is Silicon photonics. Nevertheless managing light at the nanoscale comes with many challenges, and the efficient design of nano-optical components is still an unsolved problem. In this chapter, a design methodology based on shape optimization using the adjoint method is presented.

The material here has been previously published in [43]

6.1 Introduction and motivations

Silicon photonics offer the unique ability of managing light through sub-wavelength Silicon waveguides patterned on chip, enabling extremely tight integration of photonic components and conventional CMOS electronics. Consequently, functions that previously required many separate components may now all be performed on single chips, reducing their cost, energy consumption and size [44].

Nevertheless a number of challenges remain, one of them being the efficient management of light at these scales. Indeed, while straight Silicon waveguides can have extremely low loss and enable excellent transport of light throughout the chip, other functions (such as splitters, waveguide crossings, multimode interferometers) suffer from the presence of evanescent fields outside the waveguide and imperfect reflections at the Silicon/oxide interface. These induce scattering loss, which can be highly detrimental to the total system performance. For this reason, a significant effort in photonic device topology optimization has taken place recently. This has drastically reduced the losses in Y-splitters [45, 46], crosstalk and insertion losses in waveguide crossings [46, 47], along with other more exotic components [47] and is effectively enabling better photonic circuits.

Most of these optimizations are based on heuristic optimization methods such as genetic optimization [47], particle swarm optimization [46, 48], or other hybrid methods tailored for specific problems [49]. Heuristic optimization relies on a somewhat limited parameterization of the solution space and subsequent random testing of a large number of different parameter sets. Because of the high computational cost of solving Maxwell's equations, these optimization methods may only be applied to relatively simple geometries, as they require the testing a very large number of different solutions in order to find a satisfactory one.

While this is perfectly suitable for the simple problems mentioned above, these methods will fail to perform in a reasonable amount of time for more complex geometries and functions. It is therefore necessary to have a more efficient way of performing topology optimization for general purposes. In the shape optimization approach presented here, shape derivatives play an important role. In this paper we present an adjoint method to calculate shape derivatives by wrapping an inverse algorithm around commercial Maxwell solvers. Such efficient gradient descent methods unlock the possibility to optimize particularly complex structures, which has not previously been possible.

6.2 Presentation of the adjoint method for electromagnetic problems

The adjoint method enables the computation of shape derivatives at all points in space, with only two electromagnetic simulations per iteration. It has been extensively used for shape optimization in mechanical engineering [50, 51, 52] but has seen more limited use for photonic components [53, 54, 55, 56, 57], and more recently quantum electronics [58]. Mathematical derivations of the adjoint method are available in optimization textbooks [51, 59]. Here, a very simple example that intuitively illustrates the mathematical procedure when it is used in the context of electromagnetism is presented.

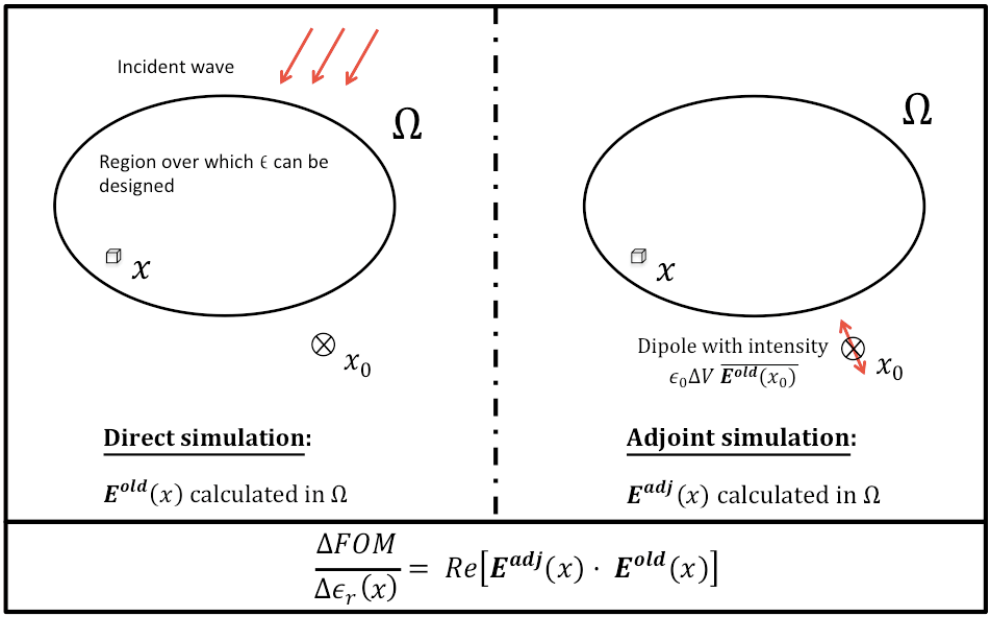

Figure 6.1: Adjoint method schematic: two simulations are needed for every iteration; the direct and the adjoint simulation. Sources for each simulation are drawn in red

In this example, the absolute value of the electric field at a given point x_0 is maximized, given a geometrical region Ω in which the electric permittivity ϵ at every point can be changed. That Figure-of-Merit is

$$FoM = |\mathbf{E}(x_0)|^2 \tag{6.1}$$

(vectors are written in bold). The change in figure of merit for a small change of dielectric permittivity $\Delta\epsilon_r$ of volume ΔV at x in Ω is

$$\Delta FoM = \Re\left[\overline{\mathbf{E}^{old}(x_0)} \cdot \Delta\mathbf{E}(x_0)\right] \tag{6.2}$$

where $\mathbf{E}^{old}(x_0)$ is the value of the electric field at a given point before any change and $\Delta\mathbf{E}(x_0)$ represents the change in electric field when the small dielectric modification is performed. Some algebraic manipulations are needed to arrive at the derivative. The change in field at x_0 can be written for a small enough volume perturbation ΔV :

$$\Delta\mathbf{E}(x_0) = \overline{\overline{\mathbf{G}^{EP}}}(x_0, x)\mathbf{p}^{ind} = \epsilon_0\Delta\epsilon_r\Delta V\overline{\overline{\mathbf{G}^{EP}}}(x_0, x)\mathbf{E}^{new}(x) \tag{6.3}$$

where $\overline{\overline{\mathbf{G}^{EP}}}$ is the Maxwell Green's function relating the electric field at x_0 to the induced polarization density $\mathbf{p}^{ind}$ at x in the infinitesimal volume ΔV. $\mathbf{E}^{new}$ is the electric field given the new dielectric distribution. If the change $\Delta\epsilon_r$ is small enough, $\mathbf{E}^{new}(x) \sim$

$\mathbf{E}^{old}(x)$ can be used as an approxiation. Noting that for binary structures $\Delta\epsilon_r$ is not small, but ΔV can be the small parameter for the derivative. A similar line of reasoning results in almost the same final equation, albeit taking care to distinguish which components of $\mathbf{E}$ and $\mathbf{D}$ are continuous across the boundary [57, 60, 61].

6.2 can be rewritten as

$$\frac{\Delta FoM}{\Delta\epsilon_r} = \epsilon_0 \Delta V \Re\Big[\overline{\mathbf{E}^{old}(x_0)} \cdot \Big(\overline{\overline{\mathbf{G}^{EP}}}(x_0, x)\mathbf{E}^{old}(x)\Big)\Big] \tag{6.4}$$

Using the reciprocity of the Green's function $\overline{\overline{\mathbf{G}^{EP}}}(x_0, x) = \overline{\overline{\mathbf{G}^{EP}}}(x, x_0)^T$:

$$\frac{\Delta FoM}{\Delta\epsilon_r} = \Re\Big[\Big(\epsilon_0 \Delta V \overline{\overline{\mathbf{G}^{EP}}}(x, x_0)\overline{\mathbf{E}^{old}(x_0)}\Big) \cdot \mathbf{E}^{old}(x)\Big] \equiv \Re[\mathbf{E}^{adj}(x) \cdot \mathbf{E}^{old}(x)] \tag{6.5}$$

The mathematical method can be understood from the new adjoint electric field:

$$\mathbf{E}^{adj}(x) = \epsilon_0 \Delta V \overline{\overline{\mathbf{G}^{EP}}}(x, x_0)\overline{\mathbf{E}^{old}(x_0)} \tag{6.6}$$

which is the electrical field induced at x from an electric dipole at x_0 driven with amplitude $\Delta V \epsilon_0 \overline{\mathbf{E}^{old}(x_0)}$, as illustrated in fig 6.1. Thus, the gradient of the Figure-of-Merit can be obtained with only a *single* simulation, even though it provides the derivative with respect to permittivity at *every* point in the computational region Ω. The term $\mathbf{E}^{old}(x_0)$ is readily available from the original forward simulation.

Therefore with just one forward simulation (which is needed to calculate the FoM in all optimization schemes) plus one adjoint simulation, the shape derivative can be obtained over the entire design region, for arbitrarily many degrees of freedom. With the gradient of the Figure-of-Merit calculated, changes in the geometry can be introduced proportional to the gradient, known as the gradient descent method. Applied iteratively, this can then lead to an optimum. For a more detailed and general study of the adjoint method and more complex Figures-of-Merit, see [60].

The adjoint method is also extremely attractive since the overall iterative scheme can be wrapped around a commercial forward solver, such as the one used in [62].

6.3 Y-Splitter optimization example using the level set method for shape representation

A Y-splitter for λ=1550nm vacuum wavelength light was optimized by the adjoint method to compare with state of the art Silicon photonic components optimized up to date[46]. The material system (Silicon waveguide, Silicon dioxide cladding) and the constraints of small

overall dimensions and minimum feature size were kept the same as in [46]. For the minimum feature size a minimum radius of curvature of 200nm was imposed. The waveguide is 220nm thick, the most common choice for Silicon photonics. The two waveguide branches and their junction at the end of the splitter were left to be the same as in [46], although they also could have easily been optimized. The design region was the central $2\mu\text{m}\times\mu\text{m}$ domain.

The method used in [46] is particle swarm optimization, which consists of calculating the Figure-of-Merit for a large population of randomly generated solutions and having the population evolve at every iteration using the information collected in the previous tests, until a satisfying solution is reached.

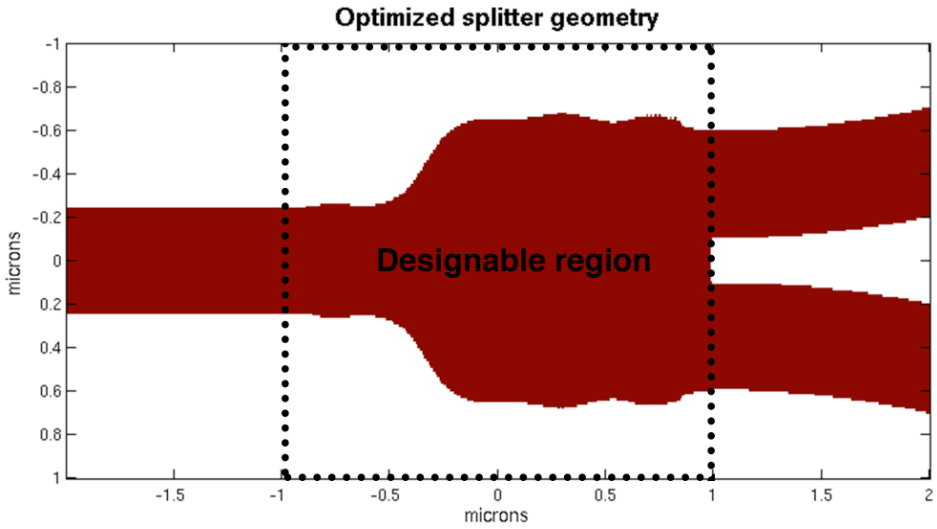

Figure 6.2: Top view of the optimized silicon splitter geometry obtained after 51 iterations of the Steepest Descent algorithm. Only the designable region geometry was allowed to change. The Silicon waveguide is 220nm thick, and the cladding is Silicon dioxide

By contrast, the adjoint method provides shape derivatives over the entire design region. The level set method, developed by Sethian and Osher [20], was chosen to represent the geometry. This enables a more flexible representation of a larger design space than, for example, spline interpolations used in [46, 47, 48]. Level sets are particularly usable within an adjoint approach, since a very large number of shape derivatives are inside the Level Set, compared to the feasible number of variables in stochastic optimization. Note also that level set methods impose two-phase, binary materials throughout the optimization, compatible with practical engineering, but in contrast with [53], which optimizes a continuously variable permittivity.

The Figure-of-Merit that we employed was transmission into the fundamental mode of

the bent output waveguides, which can be obtained from Poynting vectors:

$$FoM = \frac{1}{8}\frac{|\int \mathbf{E} \times \overline{\mathbf{H}_m} \cdot d\mathbf{S} + \int \overline{\mathbf{E}_m} \times \mathbf{H} \cdot d\mathbf{S}|^2}{\int \Re(\mathbf{E}_m \times \overline{\mathbf{H}_m}) \cdot d\mathbf{S}} \tag{6.7}$$

where $\mathbf{E}_m$ and $\mathbf{H}_m$ are the field profiles of the fundamental mode at the surface $\mathbf{S}$, while $\mathbf{E}$ and $\mathbf{H}$ are the actual fields from the direct simulation at that surface. Thus equation 6.7 is the power transmission, corrected for the mode overlap.

Adapting the adjoint equation 6.6 to the new figure of merit 6.7 (and employing an additional magnetic Green's function, $\mathbf{G}^{EM}$, and magnetic symmetries [60]), the adjoint field is:

$$\mathbf{E}^{adj}(x) = A \int \Big(\overline{\overline{\mathbf{G}^{EP}}}(x, x') \overline{\mathbf{H}_m(x')} \times \mathbf{n} - \overline{\overline{\mathbf{G}^{EM}}}(x, x') \frac{\mathbf{n} \times \overline{\mathbf{E}_m(x')}}{\mu_o} \Big) dS \tag{6.8}$$

with

$$A = \frac{1}{4}\epsilon_0 \Delta V \frac{\overline{\int \mathbf{E}^{old} \times \overline{\mathbf{H}_m} \cdot d\mathbf{S} + \int \overline{\mathbf{E}_m} \times \mathbf{H}^{old} \cdot d\mathbf{S}}}{\int \Re(\mathbf{E}_m \times \overline{\mathbf{H}_m}) \cdot d\mathbf{S}} \tag{6.9}$$

where $\overline{\overline{\mathbf{G}^{EM}}}(x, x')$ is the electromagnetic Green's function expressing the electric field at x due to a magnetic dipole at x'.

The adjoint simulation described Eq'n. 6.8 consists of sending the desired mode backwards into the splitter. This is analogous to Eq'n. 6.6, where the adjoint source was located at the measurement point of the Figure-of-Merit. This source problem can be solved with a standard Maxwell solver. FDTD is perfectly suited for this propagating wave problem. Also analogous to Eq'n 6.6 the phase of the adjoint source is set using $\mathbf{E}^{old}$ and $\mathbf{H}^{old}$, from the forward simulation, as described in Eq'n 6.9. Once the adjoint simulation is performed, the derivative of the Figure-of-Merit with respect to dielectric permittivity at *every* point in the design region is calculated by combining the forward and adjoint simulations results into Eq'n. 6.5. FDTD is perfectly suited to solve the direct and adjoint problem, which consists of propagating waves in a dielectric.

This derivative is then used to modify the geometry of the splitter. Since a level set description of geometry is employed, the derivative is used as a velocity field to modify the level set shape. This has the effect of pushing out the geometry boundary when the derivative is positive and pushing it in when it is negative. Since the refractive index if Silicon is higher than that of Silicon dioxide this implements the imperative of the derivative at every point: The Figure-of-Merit benefits from an increase in the dielectric permittivity where the derivative is positive and vice-versa. The step-size criterion for each iteration is

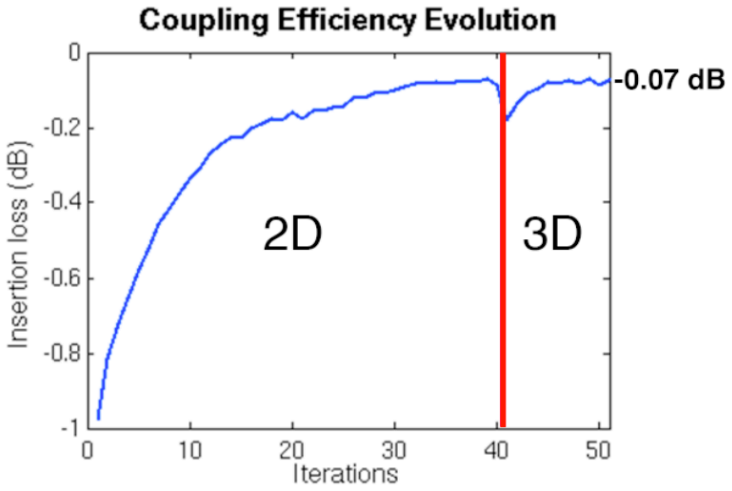

Figure 6.3: Coupling efficiency evolution during the optimization. The switch from 2d to 3d FDTD is visible at iteration 41. For comparison, the previous record of ref. [46] was -0.13dB and required 1500 simulations.

a fixed area of changing type in 2d, and a fixed volume in 3d.

The device was first optimized using 2d finite difference time domain (FDTD) simulations of a structure extruded infinitely in the 3rd dimension. In 2d, the effective index method is used and the Silicon is assigned the fictitious refractive index=2.8, which mimics the proper in-plane wavevector of the correct 3d mode. Once iterative progress stopped in 2d (41 iterations), the problem was transferred to 3d for more iterations. Naturally the first 3d iteration is not as good as the optimized 2d device, since the effective index method is only an approximation. The optimal structure was computed within 51 iterations (102 simulations), achieving a record low insertion loss -0.07dB. By comparison, ref. [3] achieved a minimal insertion loss -0.13dB, after 1500 simulations using particle swarm optimization. (Note that for such small attenuation, the simulation results are very sensitive to the simulation parameters, which may not have been perfectly identical to ref. [46]).
Thus adjoint steepest descent, with much lower computational cost, can yield as good or better results than particle swarm optimizations, which take no advantage of the underlying Maxwell equation physics.

The figure of merit evolution, as well as intermediate optimization steps, is presented in Figs. 6.3 & 6.4 respectively. There is a visible change between the 2d solution and the 3d solution, with a non-negligible efficiency improvement. This 3d improvement was only possible with the adjoint method, as the 3d computational cost limits the multiple simulations in particle swarm methods. The electric field intensity distribution of the final iteration is shown in 6.5. The large operating bandwidth of the optimized structure

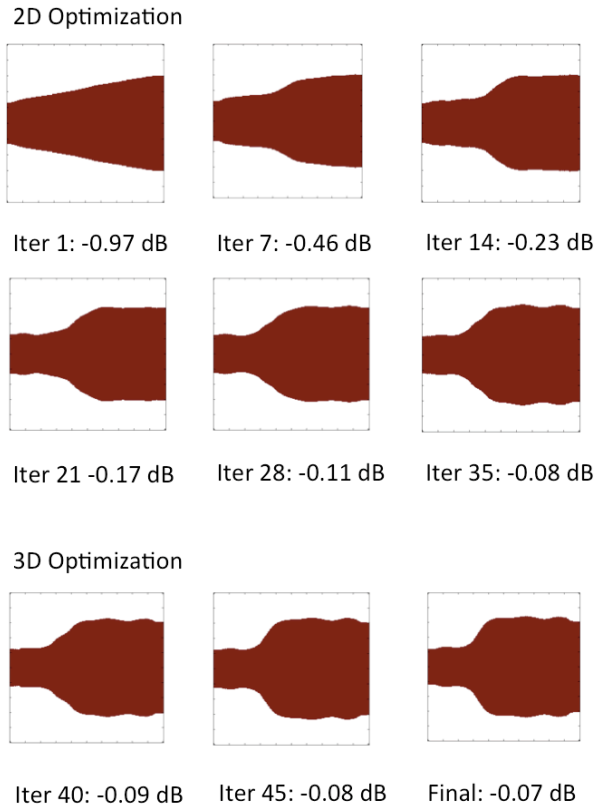

Figure 6.4: Geometry evolution during the optimization process and total coupling efficiency to the output waveguides. Iter indicates the iteration number, and the insertion loss is given in dB. The optimization is first carried out using a 2d approximation with an effective waveguide index=2.8, which mimics the 3d in-plane propagation constant. The final iterative steps are carried out in full 3d FDTD.

is shown in6.6. and is good indication of the robustness of the design generated by the optimization.

6.4 Conclusion

As photonic and wireless components become an increasingly important part of electronics, it is evident that many problems will require electromagnetic optimization. The computational cost of solving Maxwell's equations is significant, and inefficient design optimization algorithms will become unacceptable. It is shown here that the adjoint gradient decent method for shape optimization of sub-wavelength photonic devices can be readily implemented by embedding commercial Maxwell solvers within an inverse optimization algo-

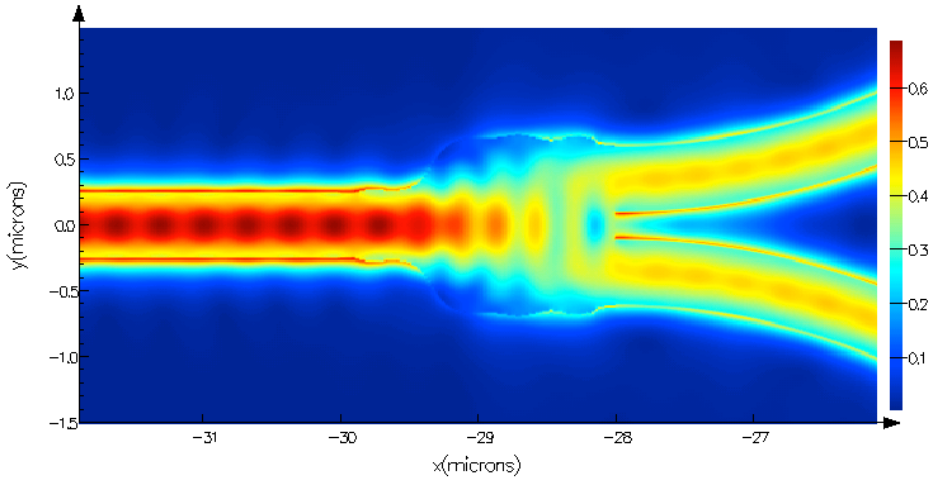

Figure 6.5: Simulated field intensity $|\mathbf{E}|^2$ for the optimized structure at λ=1550nm for a slice in the middle of the device.

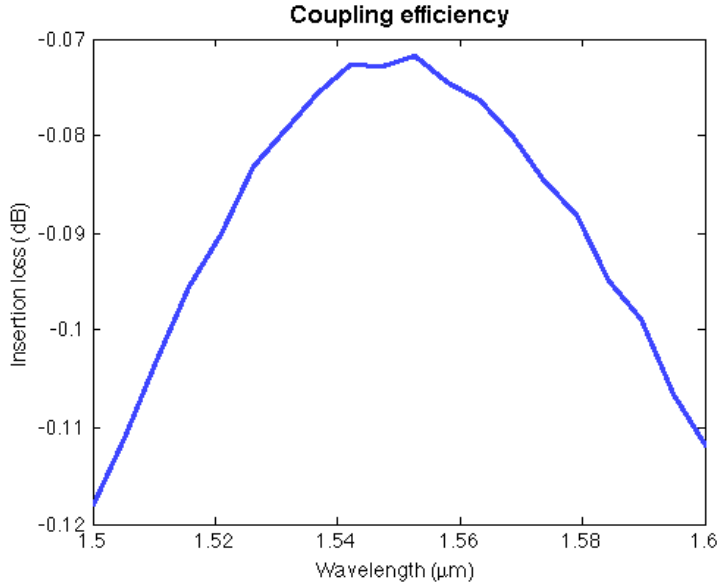

Figure 6.6: Simulated insertion loss of the optimized device for wavelengths between 1.5 and 1.6 μm. The broad operating spectrum of the device is a good indicator of the robustness of the design.

rithm.

For exploration of larger solution spaces where local optima may exist, this method may be augmented with a clever choice of Figure-of-Merit, as well as global optimization routines such as simulated annealing to provide efficient and powerful automated design of photonic components.

Adjoint-gradient-steepest-descent has already beaten the previous record for a manufacturable splitter within current Silicon photonics technology, at much less computational cost than previous methods. This opens the pathway to a more systematic, efficient, photonic component design optimization.

Appendix A

Sampler modeling and α and β factors

A.1 Alpha and Beta Factors

In the model, two factors are used to characterize the individual gain stages: $\alpha = \frac{f_a}{f_T}$, the ratio of the gain bandwidth to f_T of a replica loaded stage, and $\beta = \frac{C_{out}}{C_{in}}$, the ratio of input to output capacitance. Here we calculate α and β for simple $g_m R_L$ topology and cascode stages for the 65nm platform used.

α-factor Derivation

For a simple $g_m R_L$ topology we have

$$C_{in} = C_{ox} + AC_{gd} \tag{A.1}$$

where the second term accounts for the Miller Effect, and $C_{out} = C_{gd} + C_{ds}$. For a cascode stage, we have

$$C_{in} = C_{ox} + C_{gd} \tag{A.2}$$

Given that $C_{ox} = 0.5fF/\mu m,\ C_{gd} = 0.2fF/\mu m,\ C_{gs} = 0.27fF/\mu m$, we have $\alpha = 0.36$ for a standard $g_m R_L$ stage and $\alpha = 0.4$ for a cascode stage.

β-factor Derivation

With the expressions given above, it is easy to show that $\beta = 0.29$ for $g_m R_l$ stages and $\beta = 0.4$ for cascode stages.

A.2 Sampler Analysis

The role of the sampler is to bring the signal coming out of the amplifier to logic levels so that the digital circuit can effectively process it at the output. Most samplers rely on a positive feedback latching mechanism, such as a cross coupled inverter pair in order to achieve exponential gain and recover digital levels from extremely low signal voltages. The sampler analyzed here, and depicted in figure 4.2 is known as the StrongArm, but the presented analysis and trends can be generalized to a large family of sampler topologies, such as CML-based samplers or more exotic techniques such as double-tail sampling.

StrongArm Operating Principle

Before the sampler starts evaluating, the clock is down, and the nodes P,Q,X and Y are brought up to VDD by the reset transistors driven by clock, ϕ. The evaluation starts when the clock goes up, and is composed of two periods,: the sampling period, where in the nodes P,Q, X and Y discharge through M1, M2, M3, M4 and M7, building a differential voltage on nodes X and Y. The sampling period ends when $V_{X,Y}$ reach $V_{DD} - V_{th,P}$ and the cross coupled inverters composed of M3, M4, M5 and M6 turn on. The regeneration then starts and the differential voltage on nodes X and Y is amplified to logic level by the latch.

Sampling Period

The sampling phase can itself be divided into two separate phases. The first, during which only M1 and M2 are on, discharges nodes P and Q until they reach $V_{DD} - V_{th,N}$. The common mode voltage V_{PQ} behaves as $V_{DD} - \frac{I_1 t}{C_{PQ}}$ where $I_1 = g_{m1,2} V_{CM}$ is the current drawn by the common mode and lasts $t_1 = \frac{V_{th,N} C_{PQ}}{I_1}$

The second phase starts when M3 and M5 are also on, therefore discharging nodes X and Y. It ends when $V_{XY} = V_{DD} - V_{th,P}$. The common mode behaves according to

$$V_{XY} = V_{DD} - \frac{I_1}{C_{PQ} + C_{XY}}[(t - t_1) \\ [(t - t_1) + \tau(exp(-\frac{t - t_1}{\tau}) - 1)] \quad \text{(A.3)}$$

$$\text{where} \quad \tau = \frac{C_{XY} C_{PQ}}{g_{m,3}(C_{XY} + C_{PQ})} \quad \text{(A.4)}$$

$$\text{(A.5)}$$

There is no closed form solution to determine when nodes XY reach $V_{DD} - V_{th,P}$, but if τ is small compared to $V_{th,P}(C_{PQ} + C_{XY})/I_1$, which is usually the case, the end time of the second sampling phase may be approximated as

$$t_2 \sim \frac{V_{th,P}(C_{PQ} + C_{XY})}{I_1} + \tau + t1 \tag{A.6}$$

The differential mode, during the second phase, can be shown [12] to follow the equation:

$$\frac{d\Delta V_{XY}}{dt} = \frac{g_{m3,4}}{C_{XY}}(1 - \frac{C_{XY}}{C_{PQ}})\Delta V_{XY} - g_{m3,4}\frac{\Delta It}{C_{PQ}C_{XY}} \tag{A.7}$$

$$\Delta V_{XY}(t) = \frac{g_{m,1}}{C_{XY} - C_{PQ}}(t - t_1 + \tau_\Delta(1 - exp(\frac{t - t_1}{\tau_\Delta})) \tag{A.8}$$

$$\tau_\Delta = \frac{g_{m,3}}{C_{XY}}(1 - \frac{C_{XY}}{C_{PQ}}) \tag{A.9}$$

Since C_{XY} is usually greater than C_{PQ}, τ_Δ is usually negative, and there is no regeneration gain during the sampling period. The sampling gain can be approximated as

$$G \sim \frac{V_{thresh}}{V_{CM} - V_{thresh}}\frac{C_{PQ} + C_{XY}}{C_{XY} - C_{PQ}} \tag{A.10}$$

Regeneration Period

Once the top PMOS transistors turn on, the regeneration period starts. The approximation is made that only the cross-coupled inverter pairs are on, providing positive feedback gain, with a time constant

$$\tau_{reg} = \frac{g_{m,3} + g_{m,5}}{C_{in,D2S} + C_{out,SA}} \tag{A.11}$$

Bibliography

[1] Cisco Visual Networking Index. "The Zettabyte Era, Trends and Analysis". In: *http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.pdf* ().

[2] Christoforos Kachris and Ioannis Tomkos. "A survey on optical interconnects for data centers". In: *IEEE Communications Surveys & Tutorials* 14.4 (2012), pp. 1021–1036.

[3] Chen Sun et al. "Single-chip microprocessor that communicates directly using light". In: *Nature* 528.7583 (2015), pp. 534–538.

[4] Richard Soref. "The past, present, and future of silicon photonics". In: *IEEE Journal of selected topics in quantum electronics* 12.6 (2006), pp. 1678–1687.

[5] David Miller. "Device Requirements for Optical Interconnects to CMOS Silicon Chips". In: *Photonics in Switching*. Optical Society of America. 2010, PMB3.

[6] Behzad Razavi. *Design of integrated circuits for optical communications*. John Wiley & Sons, 2012.

[7] Edward D Palik. *Handbook of optical constants of solids*. Vol. 3. Academic press, 1998.

[8] Shun Lien Chuang. *Physics of photonic devices*. Vol. 80. John Wiley & Sons, 2012.

[9] Yimin Kang et al. "Monolithic germanium/silicon avalanche photodiodes with 340 GHz gain–bandwidth product". In: *Nature photonics* 3.1 (2009), pp. 59–63.

[10] Stewart D Personick. "Receiver design for digital fiber optic communication systems, I". In: *Bell system technical journal* 52.6 (1973), pp. 843–874.

[11] K Settaluri et al. "First Principles Optimization of Opto-Electronic Communication Links". In: *To be published* ().

[12] Behzad Razavi. "The StrongARM Latch [A Circuit for All Seasons]". In: *IEEE Solid-State Circuits Magazine* 7.2 (2015), pp. 12–17.

[13] Borivoje Nikolic et al. "Improved sense-amplifier-based flip-flop: Design and measurements". In: *IEEE Journal of Solid-State Circuits* 35.6 (2000), pp. 876–884.

[14] Jaeha Kim et al. "Simulation and analysis of random decision errors in clocked comparators". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 56.8 (2009), pp. 1844–1857.

[15] Krishna T Settaluri et al. "Demonstration of an optical chip-to-chip link in a 3D integrated electronic-photonic platform". In: *European Solid-State Circuits Conference (ESSCIRC), ESSCIRC 2015-41st*. IEEE. 2015, pp. 156–159.

[16] Erman Timurdogan et al. "An ultra low power 3D integrated intra-chip silicon electronic-photonic link". In: *Optical Fiber Communication Conference*. Optical Society of America. 2015, Th5B–8.

[17] Kwangmo Jung, Yue Lu, and Elad Alon. "Power analysis and optimization for high-speed I/O transceivers". In: *2011 IEEE 54th international Midwest symposium on circuits and systems (MWSCAS)*. IEEE. 2011, pp. 1–4.

[18] Erman Timurdogan et al. "An ultralow power athermal silicon modulator". In: *Nature communications* 5 (2014).

[19] Kengo Nozaki et al. "Photonic-crystal nano-photodetector with ultrasmall capacitance for on-chip light-to-voltage conversion without an amplifier". In: *Optica* 3.5 (2016), pp. 483–492.

[20] Jian Wang et al. "Silicon waveguide integrated germanium JFET photodetector with improved speed performance". In: *IEEE Photonics Technology Letters* 12.23 (2011), pp. 765–767.

[21] Subal Sahni et al. "Junction field-effect-transistor-based germanium photodetector on silicon-on-insulator". In: *Optics letters* 33.10 (2008), pp. 1138–1140.

[22] Ali K Okyay et al. "Silicon germanium CMOS optoelectronic switching device: Bringing light to latch". In: *IEEE Transactions on electron devices* 54.12 (2007), pp. 3252–3259.

[23] Ryan W Going et al. "Germanium gate PhotoMOSFET integrated to silicon photonics". In: *IEEE Journal of Selected Topics in Quantum Electronics* 20.4 (2014), pp. 1–7.

[24] Kah-Wee Ang et al. "Low-voltage and high-responsivity germanium bipolar phototransistor for optical detections in the near-infrared regime". In: *IEEE Electron Device Letters* 29.10 (2008), pp. 1124–1127.

[25] P Kostov et al. "PNP PIN bipolar phototransistors for high-speed applications built in a 180nm CMOS process". In: *Solid-state electronics* 74 (2012), pp. 49–57.

[26] Christopher Lalau-Keraly et al. "Ultra-sensitive detector for Silicon Photonics; a photodiode incorporating integrated bipolar gain". In: *2015 IEEE Photonics Conference (IPC)*. IEEE. 2015, pp. 429–430.

[27] Christopher Lalau-Keraly et al. "Low capacitance, high speed phototransistors with a large absorption region". In: *Energy Efficient Electronic Systems (E3S), 2015 Fourth Berkeley Symposium on*. IEEE. 2015, pp. 1–3.

[28] William Shockley, M Sparks, and GK Teal. "p- n Junction Transistors". In: *Physical Review* 83.1 (1951), p. 151.

[29] Tao Yin, Anand M Pappu, and Alyssa B Apsel. "Low-cost, high-efficiency, and high-speed SiGe phototransistors in commercial BiCMOS". In: *IEEE Photonics Technology Letters* 18.1 (2006), pp. 55–57.

[30] Raymond A Milano, P Daniel Dapkus, and Gregory E Stillman. "An analysis of the performance of heterojunction phototransistors for fiber optic communications". In: *IEEE Transactions on Electron Devices* 29.2 (1982), pp. 266–274.

[31] Joe Campbell et al. "InP/InGaAs heterojunction phototransistors". In: *IEEE Journal of Quantum Electronics* 17.2 (1981), pp. 264–269.

[32] John P Helme and Peter A Houston. "Analytical modeling of speed response of heterojunction bipolar phototransistors". In: *Journal of lightwave technology* 25.5 (2007), pp. 1247–1255.

[33] Herbert Kroemer. "Heterostructure bipolar transistors and integrated circuits". In: *Proceedings of the IEEE* 70.1 (1982), pp. 13–25.

[34] Walid Hafez, William Snodgrass, and Milton Feng. "12.5 nm base pseudomorphic heterojunction bipolar transistors achieving fT= 710GHz and fMAX= 340GHz". In: *Applied Physics Letters* 87.25 (2005), p. 252109.

[35] Michael Schroter et al. "Physical and electrical performance limits of high-speed SiGeC HBTs—Part I: Vertical scaling". In: *IEEE Transactions on Electron Devices* 58.11 (2011), pp. 3687–3696.

[36] Michael Schroter et al. "Physical and electrical performance limits of high-speed Si GeC HBTs—Part II: Lateral scaling". In: *IEEE Transactions on Electron Devices* 58.11 (2011), pp. 3697–3706.

[37] H Rücker, B Heinemann, and A Fox. "Half-terahertz sige bicmos technology". In: *Silicon Monolithic Integrated Circuits in RF Systems (SiRF), 2012 IEEE 12th Topical Meeting on*. IEEE. 2012, pp. 133–136.

[38] Peter Ashburn. *SiGe Heterojunction Bipolar Transistors*. Wiley Online Library, 2003.

[39] David Roulston. *Bipolar Semiconductor Devices*. McGraw-Hill Companies, 1989.

[40] Gianlorenzo Masini et al. "High-speed near infrared optical receivers based on Ge waveguide photodetectors integrated in a CMOS process". In: *Advances in Optical Technologies* 2008 (2008).

[41] SHINSUKE Konaka et al. "A 20-ps Si bipolar IC using advanced super self-aligned process technology with collector ion implantation". In: *IEEE Transactions on Electron Devices* 36.7 (1989), pp. 1370–1375.

[42] Sentaurus Device User Guide. "Synopsys". In: *San Jose, CA* (2008).

[43] Christopher M Lalau-Keraly et al. "Adjoint shape optimization applied to electromagnetic design". In: *Optics express* 21.18 (2013), pp. 21693–21701.

[44] Phillip Sandborn et al. "Linear frequency chirp generation employing optoelectronic feedback loop and integrated silicon photonics". In: *CLEO: Science and Innovations*. Optical Society of America. 2013, CTu2G–5.

[45] Atsushi Sakai, Tatsuhiko Fukazawa, and BABA Toshihiko. "Low loss ultra-small branches in a silicon photonic wire waveguide". In: *IEICE transactions on electronics* 85.4 (2002), pp. 1033–1038.

[46] Yi Zhang et al. "A compact and low loss Y-junction for submicron silicon waveguide". In: *Optics express* 21.1 (2013), pp. 1310–1316.

[47] Pablo Sanchis et al. "Highly efficient crossing structure for silicon-on-insulator waveguides". In: *Optics letters* 34.18 (2009), pp. 2760–2762.

[48] Yi Zhang et al. "A CMOS-compatible, low-loss, and low-crosstalk silicon waveguide crossing". In: *IEEE Photon. Technol. Lett* 25.5 (2013), pp. 422–425.

[49] Takuo Tanemura et al. "Multiple-wavelength focusing of surface plasmons with a nonperiodic nanoslit coupler". In: *Nano letters* 11.7 (2011), pp. 2693–2698.

[50] Martin Philip Bendsøe and Noboru Kikuchi. "Generating optimal topologies in structural design using a homogenization method". In: *Computer methods in applied mechanics and engineering* 71.2 (1988), pp. 197–224.

[51] Martin Philip Bendsoe and Ole Sigmund. *Topology optimization: theory, methods, and applications*. Springer Science & Business Media, 2013.

[52] Thomas Borrvall and Joakim Petersson. "Topology optimization of fluids in Stokes flow". In: *International journal for numerical methods in fluids* 41.1 (2003), pp. 77–107.

[53] Jakob Søndergaard Jensen and Ole Sigmund. "Topology optimization for nano-photonics". In: *Laser & Photonics Reviews* 5.2 (2011), pp. 308–321.

[54] Philip Seliger et al. "Optimization of aperiodic dielectric structures". In: *Journal of Applied Physics* 100.3 (2006), p. 034310.

[55] WR Frei, DA Tortorelli, and HT Johnson. "Geometry projection method for optimizing photonic nanostructures". In: *Optics letters* 32.1 (2007), pp. 77–79.

[56] Victor Liu and Shanhui Fan. "Compact bends for multi-mode photonic crystal waveguides with high transmission and suppressed modal crosstalk". In: *Optics express* 21.7 (2013), pp. 8069–8075.

[57] Georgios Veronis, Robert W Dutton, and Shanhui Fan. "Method for sensitivity analysis of photonic crystal devices". In: *Optics letters* 29.19 (2004), pp. 2288–2290.

[58] AFJ Levi and IG Rosen. "A novel formulation of the adjoint method in the optimal design of quantum electronic devices". In: *SIAM Journal on Control and Optimization* 48.5 (2010), pp. 3191–3223.

[59] Gilbert Strang. *Computational science and engineering*. Vol. 1. Wellesley-Cambridge Press Wellesley, 2007.

[60] Owen D Miller. "Photonic design: From fundamental solar cell physics to computational inverse design". In: *arXiv preprint arXiv:1308.0212* (2013).

[61] Steven G Johnson et al. "Perturbation theory for Maxwell's equations with shifting material boundaries". In: *Physical review E* 65.6 (2002), p. 066611.

[62] "Lumerical FDTD Solutions". In: *www.lumerical.com* ().