Electro-Photonic Integration and Technology Selection for Next Generation Optical Receivers

by

Vernon Elmo Paul

Department of Electrical & Computer Engineering
McGill University
Montreal Canada

September 2017

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering

Abstract

The last decade has seen an exponential increase in the magnitude and consumption of digital information. Online video streaming, cloud storage, the growing number of devices connected to the web via the “Internet of Things”, and artificial intelligence, are emerging applications that require low-latency, high-speed networks. With optical carrier frequencies operating at 100s of THz for common communication wavelengths and optical fibers having both high bandwidth and low attenuation, optical communications in data centers is imminent. However, the bottleneck lies at the interface between optics and electronics, where information is converted from one domain to the other.

This work explores historical and recent advances in improving the hardware bottleneck through “electro-photonic integration”: the dense or monolithic integration of electronic and photonic devices onto a single semiconductor die. Unlike digital circuits, where silicon is the obvious choice, the implementation of high-speed electronic and photonic hardware can be enabled by various semiconductors including silicon, silicon germanium, gallium arsenide, and indium phosphide. A comparison of these semiconductors is presented to assess their current fit for meeting future data rate demands.

Within the topic of electro-photonic integration, high-speed photodetectors are experimentally characterized and assembled with an electronic receiver. The knowledge gained from this process is then used to formulate a top-down, electro-photonic receiver design approach. The methodology incorporates three main blocks: the photodetector, electro-photonic assembly environment, and a simple, technology-based model of the transimpedance amplifier. This approach combines the otherwise separate photonic and electronic design processes in a simple, complete manner, allowing component vendors to consider co-designing or monolithically integrating devices with a technology platform that best fits customer requirements.

Sommaire

La consommation d'information numérique a connu une augmentation et une amplitude exponentielles au cours de la dernière décennie. Les nouvelles applications relatives aux services de diffusion de vidéos en ligne, au stockage infonuagique, à l'Internet des objets et à l'intelligence artificielle requièrent des réseaux à faible latence et à haute vitesse. Dès lors que les supports optiques opèrent sur des fréquences de l'ordre de la centaine de THz pour les longueurs d'ondes habituelles de communication et que la fibre optique propose à la fois une bande passante élevée et un affaiblissement faible, l'arrivée d'un système de composants communiquant entre eux par voie optique dans les centres de données semble imminente. Toutefois, l'interface entre optique et électronique, lorsque l'information est convertie d'un domaine à l'autre, conduit à un goulot d'étranglement.

Le présent mémoire propose une présentation historique des progrès réalisés pour résoudre ce goulot d'étranglement au moyen de l'« intégration électro-photonique ». C'est-à-dire par une intégration compacte ou monolithique de dispositifs électroniques et photoniques sur une même puce de semi-conducteur. Contrairement aux circuits numériques, où le silicium est un choix évident, la production de matériels électroniques et photoniques à haute vitesse peut être réalisée au moyen de différents semi-conducteurs, notamment le silicium, le silicium-germanium, l'arséniure de gallium et le phosphure d'indium. Un comparatif de ces semi-conducteurs est présenté dans le présent mémoire afin d'évaluer s'ils répondent ou non à la demande future en matière de débit de données.

Dans le cadre de l'intégration électro-photonique, les photodétecteurs à haute vitesse sont conçus et produits avec un récepteur électronique à titre expérimental. Les connaissances acquises au cours de ce processus permettent par la suite d'élaborer une approche de conception verticale pour un récepteur électro-photonique. La méthodologie comprend trois blocs principaux : le

photodétecteur, l'environnement d'assemblage électro-photonique et les matériaux, la structure et le processus de production simplifiés d'un modèle d'amplificateur d'adaptation d'impédance. Cette approche combine simplement et intégralement les processus distincts de conception électronique et photonique. À ce titre, les fournisseurs de composants pourraient ainsi envisager de produire des dispositifs selon ces deux processus ou une intégration monolithique et qui repose sur une plateforme technologique la plus adaptée aux exigences des clients.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Odile Liboiron-Ladouceur for providing me with the great opportunity to pursue graduate research work. Her strong work-ethic as well as her enthusiasm towards this thesis project and my own accomplishments as a graduate student, was invaluable.

I am grateful to Dr. Valery Tolstikhin, who along with Professor Odile Liboiron-Ladouceur, helped formulate an interest in the monolithic integration of electronic and photonic semiconductor devices. A 16-month industrial collaboration with Dr. Tolstikhin's company INTENGENT Inc., his supply of chips, modeling and analytical insights, and his breadth of experience in the field formed the main framework of this thesis.

Thank you to Professor Glenn Cowan for providing excellent mentorship and advice related to the many topics presented in this thesis. Special thanks for introducing and getting me well-acquainted with the field of integrated circuit design.

I would like to thank all my colleagues and professors from the Photonic Systems Group at McGill. Specifically, Behnam Banan and Monireh Moayedi Pour Fard, with whom I worked with extensively during experiments and theoretical discussions. Their aid was pivotal in the experimental results section of this work. A special thanks to Bahaa Radi, with whom I pursued the technology comparative study covered in Chapter 3.

Thank you to the Natural Sciences and Engineering Research Council of Canada (NSERC), who partially funded this project. Thanks also goes out to the NSERC Collaborative Research and Training Experience (CREATE) Silicon Electronic-Photonic Integrated Circuits (Si-EPIC) program for providing grants. I am grateful to Canadian Microelectronics Corporation (CMC) for providing access to computer-aided design (CAD) tools and chip packaging resources as well as to Brent Snow for providing CAD support.

I would also like thank my good friends and colleagues Xiangdong Jia and Marc Alexandre-Chan for their camaraderie and encouragement. I am grateful to Danusha, my long-time partner, for her unfailing support and inspiration during my graduate studies.

Finally, I would like to express my gratitude to my parents and family, for introducing me to the value of education and supporting me with much love and kindness throughout.

Contents

List of Figures

List of Tables

List of Acronyms

ASE Amplified Spontaneous Emission

BER Bit Error Rate

BERT Bit Error Rate Tester

BJT Bipolar Junction Transistor

CAD Computer-Aided Design

CDR Clock and Data Recovery

CMC Canadian Microelectronics Corporation

CMOS Complementary Metal-Oxide Semiconductor

CV Capacitance-Voltage

CVD Chemical Vapour Deposition

CW Continuous-Wave

DC Direct Current

DCA Digital Communication Analyzer

EPI Electro-Photonic Integration

EPIC Electro-Photonic Integrated Circuit

GCS Global Communication Semiconductors

GSSG Ground-Signal-Signal-Ground

HBT Heterojunction Bipolar Transistor

HEMT High-Electron-Mobility Transistor

IC Integrated Circuit

LC Inductor-Capacitor

LCA Lightwave Component Analyzer

LED Light-Emitting Diode

MBE Molecular Beam Epitaxy

MOCVD Metal-Organic Chemical Vapour Deposition

MOSFET Metal-Oxide-Semiconductor Field-Effect Transistor

MOVPE Metal-Organic Vapour Phase Epitaxy

NRZ Non-Return-to-Zero

NSERC Natural Sciences and Engineering Research Council

OMA Optical Modulation Amplitude

OSA Optical Sub-Assembly

PA Post-Amplification

PAM Pulse Amplitude Modulation

PCB Printed Circuit Board

PD Photodetector

PDK Process Design Kit

PRBS Pseudorandom Binary Sequence

RC Resistor-Capacitor

RF Radio Frequency

RIN Relative Intensity Noise

RMS Root Mean Square

RX Receiver

SI Semi-Insulating

SNR Signal-to-Noise Ratio

SOI Silicon-on-Insulator

TIA Transimpedance Amplifier

VCSEL Vertical-Cavity Surface-Emitting Laser

VNA Vector Network Analyzer

VOA Variable Optical Attenuator

WDM Wavelength-Division Multiplexing

Chapter 1

1 Introduction

1.1 Motivation

Information technology has been venturing into an increasing number of expanding areas in the past decade. Online video streaming services and the introduction of 4K video, high-performance computing, the internet of things (IoT), and cloud storage are just a few of many applications that demand high data rates at reduced power consumption. In order to address the ever-expanding global data traffic, communication standards are moving towards the implementation of 100G, 400G, and higher data transfer systems. Exponentially increasing data traffic poses I/O bottlenecks in data centers, which can be circumvented by optical communications [1].

In terms of speed, the bar for optical communication links is much higher than electrical links due to physical fact that carrier waves in optics operate at THz frequencies, whereas carriers in the electrical, or RF, domain operate at GHz frequencies. With copper interconnections, achieving data rates of 6 Gb/s will be limited to lengths of 75 cm [2]. Beyond these data rates or lengths, power-hungry and complex equalization techniques might have to be employed. For higher data rates and connection lengths beyond 1 m, optical interconnects offer an attractive alternative. With fiber core diameters ranging from 5 μm to 100 μm, optical links are less bulky and more flexible than high-speed electrical links that employ RF waveguides. Furthermore, the rapid advancement of silicon photonics has enabled optical waveguides with thicknesses on the order of hundreds of nanometers. These optical waveguides open the door to on-chip and chip-to-chip optical communications. The announcement of a 1.6 Tb/s optical interconnect technology by

Intel in 2013 and Corning's commercially-available, multi-purpose "Thunderbolt Optical Cables", demonstrates the rising demand for optical interconnects and its proliferation into the market.

In conjunction with the rapid growth in data traffic, the world tends towards minimizing global energy consumption. Hence it is vital for the communications industry to continually come up with innovative methods that reduce the form factor and power consumption of networking equipment in data centers and server farms. Since information storage and manipulation is primarily electronic, optical communications is possible because both electronic and photonic functions co-exist at the transmitting and receiving ends of a link. Tightly-integrated electronics and photonics, or electro-photonic integration (EPI), using silicon or III-V technology platforms are expected to deliver high-performance, low-power, and low-footprint solutions to data centers. Thus, much research is catered towards achieving high density, energy-efficient interfaces between electronic and photonic components. A system-level description of a typical optical communication link is presented in section 1.2.

1.2 Optical communication link

In Figure 1.1, the main constituents of an optical communication link, which transmits and receives digital information, is presented [3]. At the transmit end, electrical data is converted to an optical data stream with the aid of a laser driven by an electrical current. Optical information is carried through a fiber in the case of long-haul communications and in data centers, or through waveguides in the case of very short-reach, chip-to-chip or intra-chip interconnects. For simplicity, a fiber is represented in Figure 1.1. At the receive end, a photo-detector (PD) detects and converts optical data into an electrical current. As the name implies, a transimpedance amplifier (TIA) amplifies and converts this current into a voltage signal. Depending on the signal requirements of clock and data recovery (CDR) circuits, post-amplification (PA) stages (labelled "Limiter" in Figure 1.1) will typically follow the TIA to further enhance signal strength. Finally, CDR circuits recover the transmitted data. From the PD to the CDR is what comprises an optical receiver. The sensitivity of an optical receiver is a figure of merit and, arguably, the most important performance metric. It is the minimum incident optical power at the receiver required for decision circuits to recover the data with a benchmark bit error rate (BER). The bit error rate is a fraction of the number of bits that have been incorrectly decoded. Generally, a BER of 10^{-12} is desired.

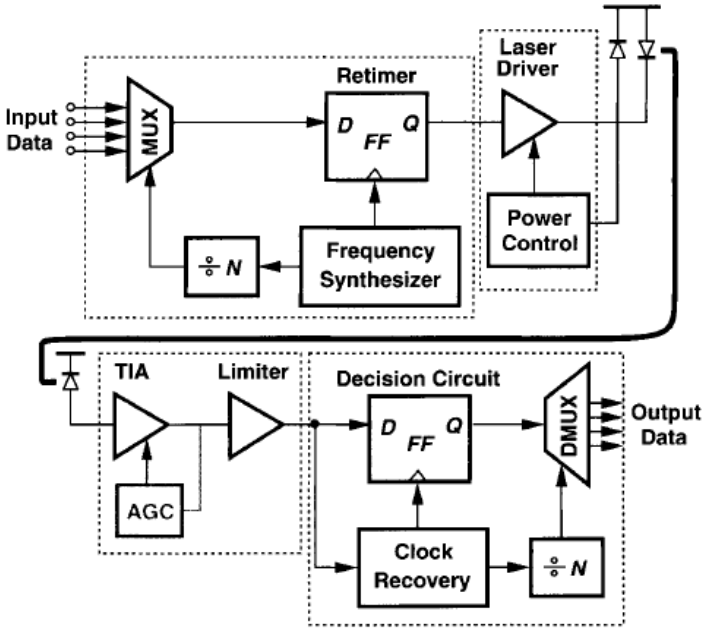

Figure 1.1 An optical communication link [3]

The optical receiver front-end is the focus of this work, which encompasses the PD and TIA. In industry, the term TIA often refers to both the initial current-to-voltage conversion stage and all amplification stages before the CDR. In all cases, the receiver front-end is simply the analog portion of the receiver. Various process technologies exist for the fabrication of the PD, TIA, and CDR circuits. For example, the PD could utilize a III-V technology platform such as InP, whereas the TIA and CDR circuits can be built in CMOS. These functional blocks can be integrated together in several ways. Traditionally, such components are designed and optimized independently before being assembled into a larger circuit. In an optical receiver front-end circuit, for example, this includes the photodetector (PD) and transimpedance amplifier (TIA), both of which are produced individually before being assembled into an electro-photonic circuit through the optical sub-assembly (OSA) stage. This is sometimes known as discrete assembly. The photonic and electronic devices remain as separate components whose electrical paths are connected via wire-bonds or flip-chip bonds. The methods of assembly are often referred to as packaging approaches. While sub-assembly has the advantage of optimizing each component's performance through selection of appropriate semiconductor platforms for fabrication, limitations emerge as bit-rates increase above 50 Gb/s. In monolithic integration, either the analog or both

analog and digital components of the receiver are produced on one chip using one process technology. Sub-assembly via wire-bonding or flip-chip bonding versus monolithic integration is described pictorially in Figure 1.2 – a major focus of this work.

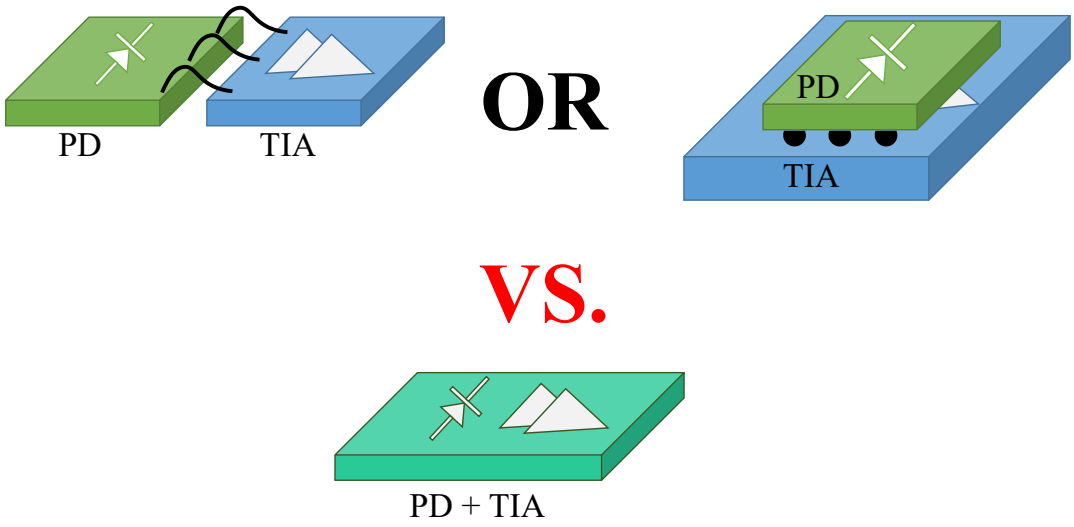

Figure 1.2 Sub-assembly (top) via wire-bonding (top left) and flip-chip (top right) versus monolithic electro-photonic integration (bottom).

1.3 Thesis objective

Spurred by the growing global abundance of digital information, the line rates in data communications are rising. With optical communications becoming the standard for meeting these demands, not every component of an optical transceiver, but the way in which they are assembled, becomes a critical design element at high speeds. By focusing on the receiver front-end of an optical communication link, this thesis provides qualitative and, more importantly, quantitative insight into the limitations of optical sub-assembly (OSA) in this high-speed space. To address these limitations, current and future semiconductor technology platforms and packaging solutions are discussed. Specifically, the system benefits that can be attained from the monolithic integration of electronic and photonic components on a single die are outlined. This is facilitated by using empirical evidence from literature, theoretical reasoning, experimental characterization, and validations based on simulation models. Valuable technology-based comparative studies exist for wireless applications, such as in [4] and [5], where the compared metrics are significant for RF

transceivers. In this study, the comparison is done in the context of optical transceivers, where carrier frequencies are higher and circuit architectures are different. An emphasis placed on the classical fundamentals of receiver design and analysis, as it aids the communications community understand how the demands of today contrast those in the early days of long-haul fiber communications, and the implications it has on the technology and assembly selection of future circuits. Performance is assessed using the following key metrics: low power, low footprint, and low complexity. Receiver sensitivity is directly related to the power consumed by an optical link, as it establishes the link budget of the system. When analyzing and designing front-end circuits, it is a metric of paramount importance. These criteria are selected because they have implications in lower costs for both data centre operators and transceiver vendors.

1.4 Thesis organization

An overview of integrated electro-photonic solutions spanning CMOS, SiGe, and InP technologies is presented in Chapter 2. In Chapter 3, system-level receiver theory is presented and followed by a review of the state-of-the-art in receiver front-end design. In Chapter 4, the experimental procedure and results of testing high-speed photodetectors is summarized. Combining the knowledge from Chapters 3 and 4, Chapter 5 goes through a top-down design methodology of optical receiver front-ends, highlighting the advantages of electro-photonic co-design and monolithic integration of the photodetector and rest of the optical receiver front-end.

Chapter 2

2 Electro-photonic integration

Monolithic integration of electronic and photonic devices in the broad field of optical communications presents an excellent opportunity for compact receivers and transmitters. If designed well, the signal integrity of integrated devices can be of much higher quality than hybrid devices, where multiple chips are interconnected via a board or flip-chip techniques [6]. Packaging and bonding parasitics are detrimental to performance and add complexities during the final integrated circuit design stages where careful modeling of parasitics must be considered [7]. The inevitable caveat with higher degrees of monolithic integration is the limited ability to optimize each device's performance independently. For example, the fabrication processes used for electronics may not be optimal for photonics. However, the elimination of electrical parasitics may provide more room for device optimization, which is the core of this work.

Thus, there are many compelling reasons for pursuing monolithically-integrated solutions. Often, the primary question in the process of integration concerns the selection of the substrate material. Group IV semiconductors such as silicon or those based on both Group III and V (III-V or compound semiconductors) are common substrate choices. These are simply substrate materials, upon which various other alloys, compounds, and metals can form epitaxial layers for different purposes. For example, InGaAs is a common epitaxial layer grown on InP substrates for light emission in laser diodes and detection in photodiodes [8]. In this chapter, the basic properties of silicon and III-V semiconductors as well as photodetectors are covered. Using this information,

monolithic electro-photonic integration strategies that exist in each semiconductor platform are discussed.

2.1 Semiconductors

An important feature and comparison metric between III-V semiconductors and silicon semiconductors is their respective bandgap structures. In silicon materials, a graphical plot of the conduction and valence energy bands as a function of momentum, or the wave vector *k*, yields an offset between the minimum of the conduction energy band and the maximum of the valence energy band in the *k* direction. It is this specific characteristic that renders certain semiconductors, such as silicon, as having an "indirect bandgap" as the bandgap is defined between valence band maximum and conduction band minimum. In compound, or III-V semiconductors, the minimum and maximum of the conduction and valence bands lie at the same momentum. Therefore, processes such as the creation and annihilation of electron-hole pairs can occur without any changes in momentum – a far more efficient process [9].

The direct bandgap structure of III-V semiconductors has been instrumentally exploited in the creation of a vast number of photonic devices including but not limited to solar cells, light-emitting diodes (LEDs), laser diodes, quantum dot lasers, and photodetectors. The bandgap of III-V semiconductors can vary to within the visible light region and spans portions of the infrared region. III-V semiconductors have indeed enjoyed a long history of success and prominence in photonics, especially optical communications, where communication bands lie at various portions of the infrared spectrum [10].

The energy bandgap of GaAs and InP are 1.42 eV and 1.35 eV, which corresponds to photon wavelengths of 873 and 918 nm, respectively [11]. The GaAs bandgap lies just above the multi-mode fibre optics communication wavelength of 850 nm, which is also the wavelength at which Vertical Cavity Surface Emitting Lasers (VCSELs) are optimized to emit [12]. Although the bandgaps of InP and GaAs are below common communication wavelengths such as 1310 nm and 1550 nm, ternary alloys such InGaAs or quarternary alloys such as InAlGaAs grown on these III-V substrates allow for longer-wavelength emission and detection [13]. Metalorganic vapor-phase epitaxial (MOVPE) reactors, molecular beam-epitaxy (MBE) systems, and metalorganic chemical vapour deposition (MOCVD) are fabrication technologies that have enabled the growth

of the aforementioned III-V based alloys in the clean room [13]. For reference, InGaAs properties are also included in Table 2.1, which lists semiconductor properties relevant to high-performance electro-photonic devices. Germanium, which is often grown on silicon or combined with it to form SiGe, is also listed in Table 2.1 because it is another enabling semiconductor for light detection and high-speed electronics [11, 14, 15]. InGaAs is a direct bandgap semiconductor whereas Ge (like Si) is an indirect bandgap semiconductor. Note that both InGaAs and Ge bandgaps correspond to wavelengths above the 1550 nm communication band, which is why they are excellent absorbers of higher energy photons of 1310 and 1550 nm wavelengths. Conversion from energy E_{ph} to photon wavelength λ is done by equation 2.1, where c is the speed of light and h is Planck's constant. Photons with energies less than the bandgap of a crystal will not be absorbed, whereas those with energies right above the bandgap will be absorbed.

$$E_{ph} = \frac{hc}{\lambda} \tag{2.1}$$

Another useful comparison metric between different types of semiconductors are their electron and hole mobility values, which plays a large part in how efficiently and quickly charge carriers such as electrons and holes can get from one location to the next in the presence of an electric field. Of course, this is very important in both purely electronic devices such as transistors and optoelectronic devices such as photodetectors. The average or drift velocity v_d of a carrier in the presence of an electric field E is related to its mobility μ via equation 2.2.

$$v_d = \mu E \tag{2.2}$$

In this domain, III-V semiconductors such GaAs and InP have much higher electron mobility values when compared to silicon. The hole mobility values of InP and GaAs are less than silicon, but to a lesser extent. As will be discussed later, III-V electronic and optoelectronic devices have often been designed in such a way that enables electrons to be the main carrier that governs device performance.

Table 2.1 Properties of commonly-used opto-electronic semiconductors [9], [14]–[16]

Semiconductor	Bandgap (eV)	Bandgap (λ)	Electron Mobility ($cm^2\ V^{-1}\ s^{-1}$)	Hole Mobility ($cm^2\ V^{-1}\ s^{-1}$)
Si	1.11	1127 nm	1450	490
Ge	0.66	1879 nm	3900	1900
GaAs	1.42	873 nm	8500	400
InGaAs	0.75	1653 nm	10000	250
InP	1.35	918 nm	4600	150

2.2 Photodetectors

The photodetector, which is a semiconductor device, converts a received optical power into an electrical current. Optical power is composed of discrete packets of energy called photons. Each photon has an energy proportional to its wavelength, as shown previously in equation 2.1.

When a photon of finite energy is incident upon a semiconductor, an electron-hole pair is created. An electron-hole pair is also known as a charge carrier because it is available for charge conduction. The most fundamental configuration of a photodetector, that facilitates optical-electrical conversion, is a reverse-biased semiconducting diode of n and p doping regions. An electron-hole pair generated within the depletion region of the diode is exposed to an electric field and can drift, and is thus involved in the generation of an electric current known as drift current. Charge carriers generated outside the diffusion region must diffuse to the depletion region and be swept across to contribute to an electric current. This is known as diffusion current and is a slower process compared to drift current. If the majority of carriers are generated within the depletion region, the drift current should be dominant.

Both carrier diffusion and/or transit through the depletion region is quantified by its transit time τ_R. If the transit time is relatively long, the speed of the photodetector will be limited by this process, i.e. transit-time limited. If the detector is designed to enable fast transit times, then the speed may be limited by the parasitic capacitance C_{PD} and resistance R_{PD} that results from depletion region and metallic contacts, i.e. RC limited. The bandwidth (BW) of a photodetector can be described by combining these two effects as in (2.3) [17].

$$BW = \frac{1}{2\pi}\frac{1}{\tau_R + R_{PD}C_{PD}} \tag{2.3}$$

If photodetector current is dominated by drift current, then the τ_R in equation 2.3 can be expressed in terms of the width W or thickness of the depletion region and carrier drift velocity as shown in equation 2.4 [17]. The drift velocity is related to the carrier mobility and strength of the electric field in the depletion region, as described earlier in equation 2.2.

$$BW = \frac{1}{2\pi}\frac{1}{W/v_d + R_{PD}C_{PD}} \tag{2.4}$$

In equation 2.4, decreasing the width of the depletion would reduce the transit time and potentially improved the bandwidth of the photodetector. Of course, this may also reduce the number of carriers generated in the depletion region. Thus, the speed and efficiency of a photodetector is a common design trade-off.

The quantum efficiency η of a photodetector is the ratio of charge carriers that contribute to photo-current to the number of photons incident on the detector. The quantum efficiency is a property of the materials used in the detector, the physical design and electric field within the absorption regions of the detector, and the absorption properties of the detector for a given wavelength. Responsivity, expressed in A/W, is the current per incident optical power. It is directly proportional to the quantum efficiency as defined below [18], where q is the charge of an electron:

$$R = \frac{\eta q \lambda}{hc} \tag{2.5}$$

The sensitivity of optical receiver front-ends are directly proportional to the responsivity of the photodetector, which is why the responsivity is a crucial parameter that is often optimized by photonics designers. Responsivity can be improved by increasing the absorption or depletion region of the detector so that more photons are converted to charge carriers. However, as demonstrated in equations 2.3 and 2.4, this has a consequence on the realizable bandwidth of the detector either through increased parasitics R_{PD} and C_{PD} or increased transit time of carriers. This is a common challenge in designing high-speed photodetectors.

2.3 Transimpedance amplifiers

In the front-end receiver chain presented in Figure 1.1 and Figure 1.2, the transimpedance amplifier (TIA) or electronic portion of the receiver follows the photodetector. Its role is to convert the photo-current I_{PD} generated by the photodetector into a voltage V_{OUT} with a transimpedance gain Z_t, as described in Figure 2.1. The transimpedance gain can be expressed in Ω or in dBΩ if $20\log_{10}(|Z_t|)$ is taken. Post-amplification stages, digital logic, passive components like resistors and capacitors, and feedback systems are also incorporated on the same die as the TIA stage, forming an integrated circuit (IC). As mentioned in chapter 1, the complete IC described above is sometimes referred to as the TIA for simplicity. All, or a combination of, the semiconductors mentioned in Table 2.1 have been utilized to design and manufacture TIAs. The critical electronic device that provides amplification and logic in these ICs is the transistor, of which there are many variants, with each offering its own set of benefits and shortcomings.

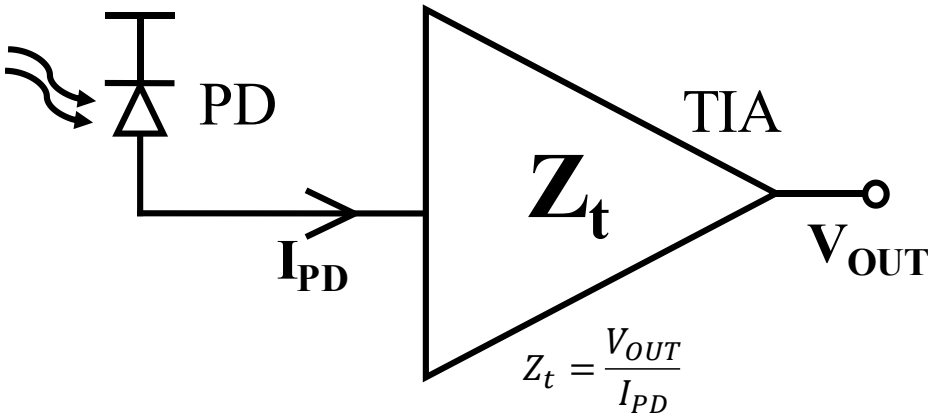

Figure 2.1 System-level schematic of an optical receiver front-end with the TIA providing transimpedance gain Z_t.

2.3.1 Transistor technologies

When using silicon, the most common type of transistor is the metal-oxide-semiconductor field-effect transistor (MOSFET). The three main terminals of the transistor are the gate, source, and drain, depicted by *G*, *S*, and *D* in Figure 2.2, respectively. There is a very insulating layer,

such as an oxide, between the gate and the body of the transistor. Depending on the polarity of the voltage applied to the gate, a p-channel (for PMOS transistors) of conducting holes or an n-channel (for NMOS transistors) of conducting electrons are formed near the interface of the insulator and semiconductor [19]. Then, current flows between the source and drain terminals of the transistor. This current is known as the drain-source or simply, drain current. Typically, p-type silicon is used as the substrate material. IC designs that implement MOSFETs will usually exploit the mature and well-known complimentary metal-oxide-semiconductor (CMOS) process, where both p-channel (PMOS) and n-channel (NMOS) devices are utilized. The schematic representation of NMOS and PMOS transistors is provided in Figure 2.2.

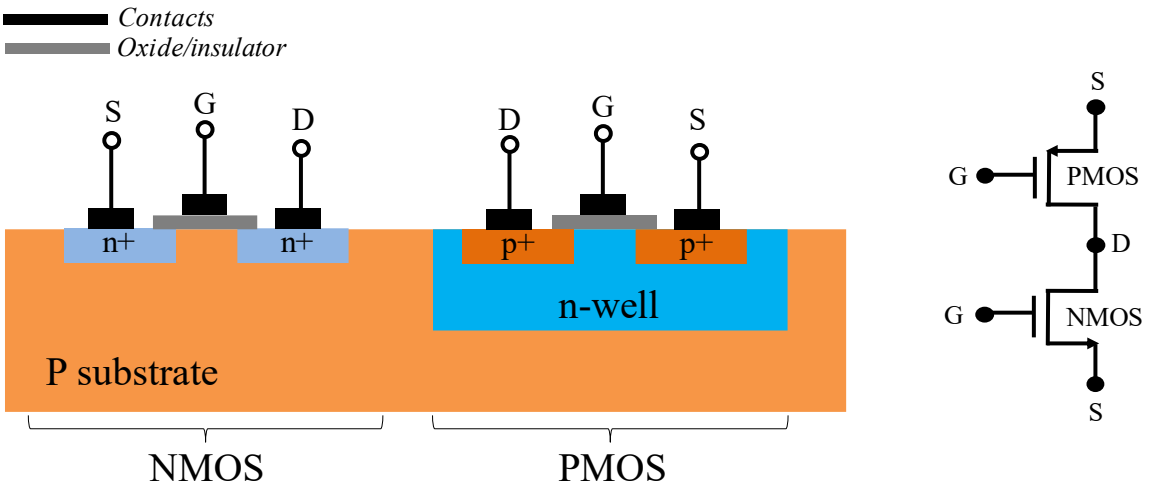

Figure 2.2 Cross-section (left) of transistors built using a CMOS process. The equivalent electrical schematic is shown on the right.

CMOS technology enables densely-integrated, low-power digital logic and is one of many reasons for its dominance in the digital consumer market [20]. The feature size or gate length of a MOSFET in CMOS is a critical feature, which has been scaling down continuously over many decades due to sustained success in fabrication technology. This trend has been enabling digital ICs to be more dense, powerful, and less costly with the introduction of each new gate length. This trend is known as technology scaling and new each technology node offers higher-performing and denser ICs [21]. For example, the scaling of CMOS devices improves the speed of transistors.

Thus, CMOS is also proliferating into high-speed analog IC markets including wireless and optical communications [5]. As a well-established and widely-used technology with increasing performance gains brought by each new technology node, TIAs built in CMOS is a competitive platform for next-generation optical communication products. The CMOS process commences with the creation of a silicon wafer, which explains why the terms silicon and CMOS are often used interchangeably in the semiconductor industry. In section 2.5, an overview of photonic devices built in silicon will be provided.

The bipolar junction transistor (BJT), illustrated in Figure 2.3 is an older technology that typically offers higher transconductance (voltage to current conversion gain) and speed than MOSFETs [21]. Thus, BJTs are still widely used in microwave and RF circuits such as power amplifiers. As optical communications are moving towards 400G and 1600G standards, the use of BJTs in TIAs is a reasonable choice. BJTs can be made in Si, GaAs, or InP, with GaAs and InP BJTs providing higher speeds than Si BJTs due to higher carrier mobilities.

Like the MOSFET, the BJT is a three-terminal device consisting of an emitter, base, and collector of n, p, and n semiconductor doping profiles, respectively. In the n-p-n format, electrons from the emitter are injected into the base, diffuse, and get swept across the collector, forming collector current [19]. The p-n-p format is also available, but is slower than n-p-n transistors because the collector current is formed by holes instead of electrons.

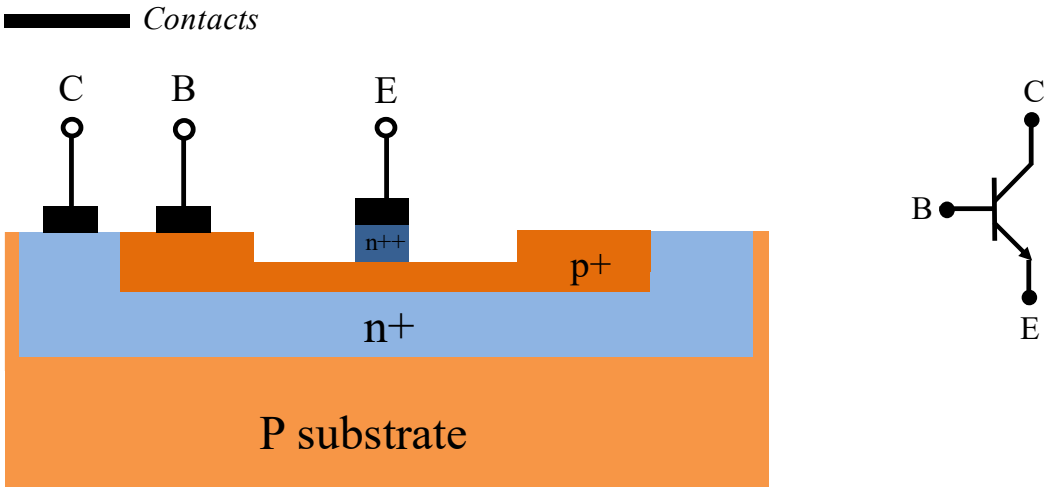

Figure 2.3 Cross-section (left) of an npn BJT. The equivalent electrical schematic is shown on the right.

The more common form of the BJT that is used in practical, high-speed applications is the heterojunction bipolar transistor (HBT) [22]. The HBT will implement a lower bandgap material or semiconductor alloy of varying concentration in the base region to induce a higher drift field, which leads to a higher carrier drift velocity [23]. This is also known as "bandgap engineering" [13]. Hence, electrons injected from the emitter will cross the base region and into the collector with faster transit times. For this reason, HBTs are capable of handling higher signal speeds than BJTs. InP HBTs will use an InGaAs base. The base of SiGe HBTs will have a graded concentration of Ge in Si. Historically, bipolar devices (either BJTs or HBTs) can operate at higher transconductance and speed than MOSFETs [19]. However, due to the scaling of MOSFETs, it is important to assess the current state-of-the-art of all these technologies in the design of high-speed TIAs. This assessment is provided in chapter 3.

Since SiGe HBTs use a silicon substrate, it is compatible with the CMOS process. In fact, the combination of SiGe HBTs with CMOS devices using a single process technology is known as the SiGe BiCMOS process technology [24]. It combines the best of both technologies and offers HBTs for high-speed capabilities and MOSFETs for digital logic capabilities. TIAs built in this process are also included in the literature survey that will be presented in the next chapter. Since Ge is a better photon absorber than Si at wavelengths of 1310 and 1550 nm, the SiGe BiCMOS process offers electro-photonic integration possibilities at these wavelengths.

A summary of the three technology platforms discussed in this thesis are summarized in Table 2.2, where the secondary semiconductors form critical layers of devices. Note that each of these technologies advance on a consistent basis. In CMOS, each new technology node offers smaller transistor dimensions, lower power consumption, and higher performance. For HBTs built using InP or SiGe BiCMOS process, each technology will also scale down transistor dimensions or offer higher performance through better growth of semiconductor alloys and bandgap engineering techniques.

Table 2.2 Summary of optical receiver semiconductor technologies

Process	Substrate Semiconductor	Secondary Semiconductors	High-speed transistors	Logic transistors
InP HBT	InP	InGaAs	HBT	HBT
SiGe BiCMOS	Si	SiGe	HBT	MOSFET
CMOS	Si	-	MOSFET	MOSFET

2.3.2 Circuit topologies

The role of the transimpedance amplifier (TIA) is to convert signal current generated by the photodetector into a voltage with as much transimpedance gain, little noise, and high bandwidth. Figure 2.1 was a simple block diagram of the PD and TIA, but a somewhat more detailed diagram such as Figure 2.4 is required to understand the important design requirements of a TIA.

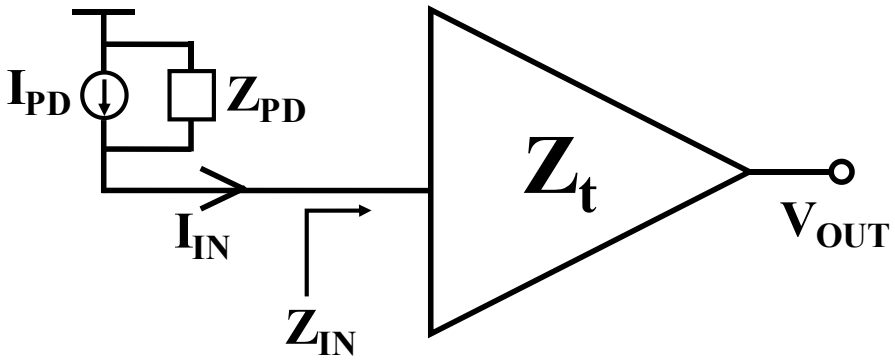

Figure 2.4 Common electrical representation of PD-TIA interface with the photo-current generated by PD represented as a current source with source impedance Z_{PD}.

Electrically, the PD is represented as a current source with an impedance (mostly capacitive) Z_{PD}, while the TIA will present its own impedance (a combination of capacitive and resistive) Z_{IN}. A more detailed characterization and analysis of these impedances will be presented in chapters 3 and 4. Figure 2.4 is essentially a current divider. It is therefore important to minimize

Z_{IN} relative to Z_{PD} so that more I_{IN} can be injected into the TIA for amplification. As the signal frequency of I_{PD} increases, the capacitive components of Z_{PD} will reduce Z_{PD} relative to Z_{IN}, minimizing I_{IN} and, hence, the bandwidth. Thus, the TIA designer must minimize the input impedance for better bandwidth. However, the mechanisms that increase bandwidth also lower transimpedance gain and/or increase power consumption. Larger transimpedance gain is also favourable for reducing noise and increasing sensitivity. These are common trade-offs that make design of high-speed TIAs challenging.

The combination of Z_{PD} and Z_{IN} at the input of the TIA is naturally very large as it composed of PD and PD-to-TIA connection parasitics. As such, the overall bandwidth of the receiver system is heavily governed by the input environment of the TIA [17]. In the simplest case, the PD-TIA interface can be modelled as a parallel resistor-capacitor ($R_{in}C_{in}$) combination at the input. The C_{in} is a total capacitance that includes PD's junction, stray capacitance from traces or bond-wires to ground, bond-pad capacitance, etc. The R_{in} is typically the input resistance of the TIA since this is the smallest parallel resistor in a PD-TIA interface. This forms an input pole ωp_{in}, in units of rad/s, which contributes to the frequency and time-response of the system. Inductive components such as bond-wires will be explored in greater depth in chapter 5.

For a first order system that has a single pole, the pole is the frequency at which the magnitude-squared of the system transfer function reaches half of its DC value. In other words, the pole coincides with the 3-dB bandwidth of system. Higher order systems have multiple poles but the overall bandwidth is limited by the lowest-frequency (or in other words, dominant) pole. Of course, an optical receiver will have many poles but the electro-photonic designer will try to places these poles as far as possible from ωp_{in}, so that the overall bandwidth of the system is minimally impacted by them. As such, ωp_{in} is often the dominant pole in an optical receiver. In cases where PD and TIA input parasitics are minimized (i.e., monolithic integration), the ωp_{in} may not be the dominant pole. This is not necessarily a performance-limiting characteristic because ωp_{in} will be larger (higher frequency) in this case; enabling higher bandwidth. However, there is usually a responsivity trade-off associated with minimizing photodetector junction capacitance by making it smaller. This trade-off is described in in the experimental characterization of a high-speed photodetector in chapter 4.

There are two primary TIA configurations: open-loop and closed-loop. In the open-loop configuration, the TIA in Figure 2.4 will be implemented with an amplifier topology that directly

takes the photo-current I_{IN} and amplifies it to a voltage V_{OUT} via Z_t. In the closed-loop configuration, the amplifier block Z_t will consist of a voltage-to-voltage amplifier with open-loop gain A and a negative feedback around it to provide a Z_t. Usually, the negative feedback element is a resistor and configurations of this type are more commonly known as shunt-shunt feedback TIAs [25]. The mathematical relationships that describes the operation of closed-loop TIAs will be described in chapter 5.

The advantage of closed-loop configurations is that the negative feedback reduces the amplifier's input resistance. This helps increase the bandwidth of the TIA and provides the designer with the flexibility of choosing from a wide variety of amplifier topologies (both low and high input impedance topologies). The choice of the amplifier in the closed-loop configuration can either be a common source, common source with degeneration, or inverter-based amplifier for MOSFETs and a common emitter or common emitter with degeneration for bipolar devices. In all the closed-loop amplifier topologies, the signal at the output must be inverted with respect to the input for negative feedback. Note that the common source and common emitter configurations both have this inverting property as well as a fairly a high input impedance, but this is reduced by the feedback network.

In the open-loop configuration, the designer is limited to selecting amplifiers with low input impedance because feedback does not reduce the input resistance. An amplifier topology that provides low input impedance and high gain is the common gate and common base topology for MOSFET and bipolar devices, respectively.

In the next chapter, the transconductance g_m of transistors will be discussed further. Still, a summary of open-loop and closed-loop configurations and their respective, common amplifier topologies for both MOSFET and bipolar devices is provided in Figure 2.5. Since the g_m of bipolar devices is generally much higher than MOSFET devices, linearity can be a concern. Thus, a common-emitter with emitter R_E degeneration is shown because this topology improves linearity as the open-loop gain becomes almost exclusively dependent on the ratio R_C/R_E. While the feedback resistor R_f provides transimpedance gain Z_t in the closed-loop configuration, the collector and drain resistors R_C and R_D provide Z_t in the open-loop configuration.

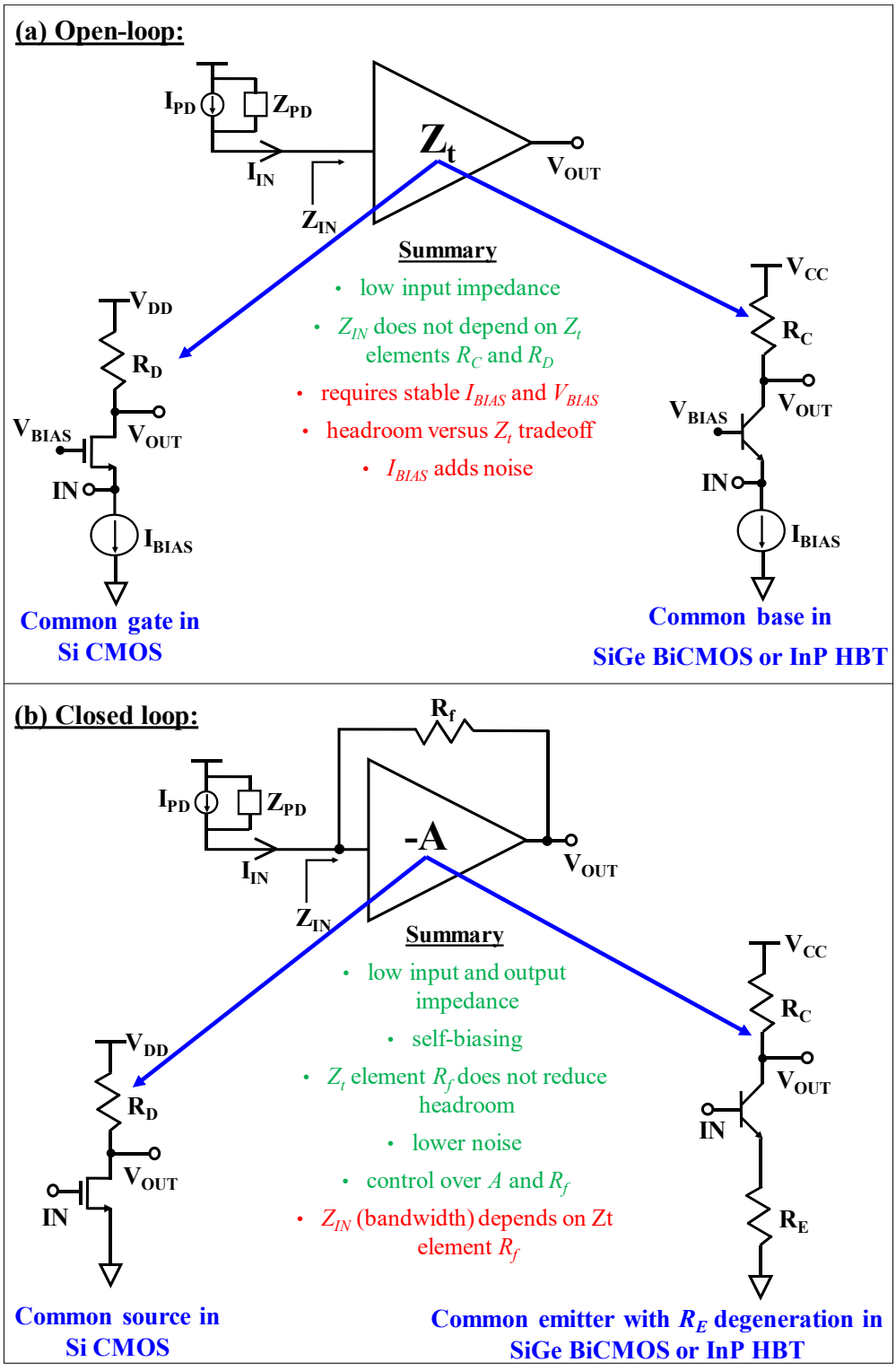

Figure 2.5 Summary of common open-loop and closed-loop amplifier topologies using both MOSFETs and HBTs

2.4 III-V electro-photonics

Indium phosphide (InP) is a popular material in the high-speed space of photonic and electronic devices. All components of an optical communication link, as described in Chapter 1, have been successfully implemented using an InP substrate, including tunable lasers, Mach-Zender modulators, and coherent receivers [8, 13]. In the case of photodetectors, well-known choices for absorption region alloys include GaAs and In(x)Ga(1-x)As.

The most commonly explored method of integration using III-V substrates is monolithic electro-photonic integration of the receiver front-end. Both academia and industry have successfully demonstrated the integration of a photodetector and TIA built using heterojunction bipolar transistor (HBTs) [6, 17,18].

A cross section of an InP-based monolithic electro-photonic front-end circuit is presented in Figure 2.6 [6]. The figure appeared in a 1996 IEEE International Electronic Devices Meeting paper by Bell Labs. The Indium Phosphide substrate is doped with Iron and the epitaxial layers are grown atop. On the right-hand side, an equivalent electrical schematic is drawn, where the interconnect between the PD and TIA is shown in red.

In this figure, the cross-sections of a p-i-n photodetector with an npn HBT transistor is depicted. The light is shown incident from the bottom of the wafer. The input light can either travel directly through the substrate or by means of a through-wafer via created in post-processing. One advantage of this type of illumination is that the light avoids interacting with the metallic interconnect layers above, which are detrimental to the responsivity of the detector. In the case of a common emitter TIA configuration with feedback, there would be an electrical path routed (in red) between the base *B* of the HBT and p+ contact of the detector. Not shown in the electrical schematic is any feedback path that may exist between the input *B* and output *C* of the common emitter TIA.

The integration of front-end receivers with a photodetector in III-V, specifically InP, dates to the late 1980s. In Figure 2.6 it is evident that the HBT and p-i-n photodetector devices share similar cross-sections. In fact, the first three layers above the InP substrate consist of the same alloys and relative doping levels. To clarify, the n- lightly doped InGaAs layer of photodetector in Figure 2.6 plays the role of the intrinsic region, where the strongest electric field exists for the facilitation of electron-hole pair drift currents [28].

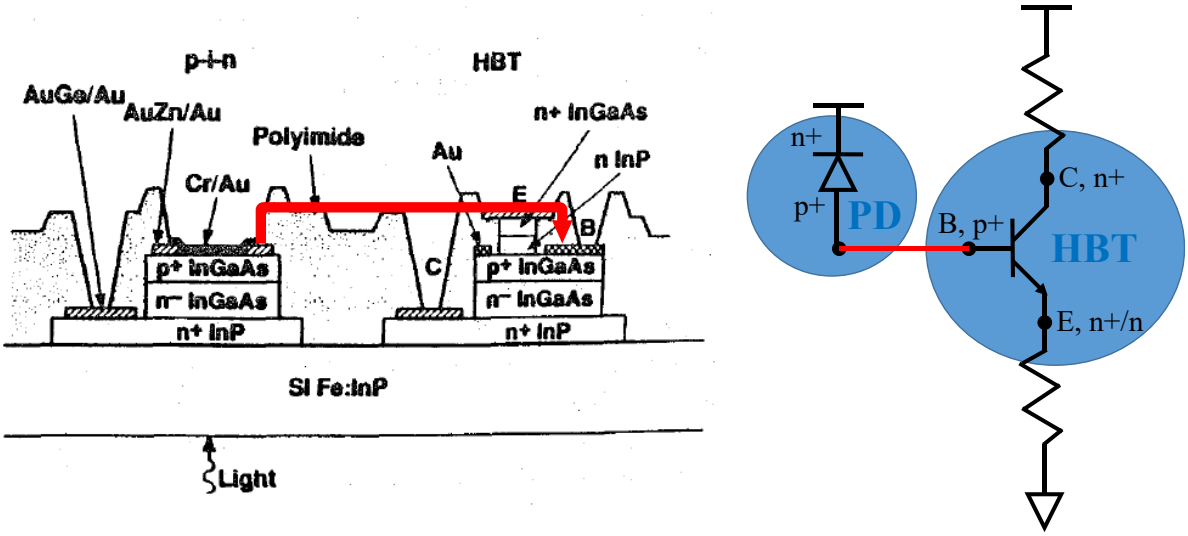

Figure 2.6 Cross-section (left) of an InP p-i-n photodiode monolithically integrated with an HBT-based TIA by Bell Labs [6]. Equivalent electrical model (right) of configuration.

The group at Bell Labs reported a 3-dB bandwidth of 10.4 GHz with a total power dissipation of 92 mW. In the year 2000, a monolithically-integrated InGaAs-InP pin/HBT was demonstrated to have a small-signal bandwidth of 30 GHz, and a reasonable eye opening at 40 Gb/s [27].

Despite these advancements, recent progress in the field of III-V monolithic integration has slowed down. This could be due to several reasons. One, is that photodetector or TIA performance is often sacrificed when they are implemented using a single fabrication process. Optical sub-assembly via wire-bonding is still widely employed in industry and works for 25 Gb/s. The continued dominance of the CMOS process in digital electronics, which has high yield and higher integration densities, spurred both academic and industry interest in the field of "silicon photonics". As will be discussed in the following section, this is a competing electro-photonic integration technology. Pure photonic components such as waveguides, photodetectors, and electro-optic (E/O) modulators have been built using both InP and Si semiconductor platforms. As the field of silicon photonics advances and single-lane data rates rise, this thesis evaluates the possibility of reviving some of the III-V electro-photonic integration strategies discussed above. Perhaps higher data rates will demand the minimal parasitics imposed by monolithic integration.

2.5 SiGe and Si electro-photonics

The semiconductor properties of silicon, including its indirect bandgap and absorption properties at communication wavelengths, imposes challenges in realizing true monolithic electro-photonic circuits. For silicon to achieve the same amount of absorption as GaAs at a wavelength of 850 nm, it would have to be about 20 times thicker [7]. This physical reality contradicts the thin-film nature of CMOS processes. The longer absorption length of silicon also puts a constraint on the achievable bandwidth of the photodetector. This is due to either large junction capacitance if the bandwidth is RC limited or a longer carrier transit time if it is transit-time limited.

One common approach to overcoming the limitations of silicon, is to implement Ge into the CMOS process. Ge crystals have excellent responsivity from visible to near-infrared wavelengths and have faster electron and hole mobilities when compared to silicon. The advancement in the field of Ge integration was driven by the development of a process that allowed the epitaxial growth of germanium on silicon, despite the 4.2 % lattice mismatch between the materials. SiGe buffer layers and annealing processes are some of the ways in which dislocation densities were reduced [29].

The marriage between silicon and germanium as material systems motivated the inclusion of both pure silicon electro-photonic integration and germanium-based electro-photonic integration in this section. However, a distinction must be made. The SiGe BiCMOS process is a hybrid process that benefits from the use of both BJTs and MOSFETs. The BJTs or HBTs are implemented using the semiconductor alloy SiGe, not a thin epitaxial layer of germanium grown on silicon. The MOSFETs are implemented using the CMOS process. The substrate is usually p-doped silicon. For an optical receiver front-end built in SiGe BiCMOS, the analog portion will exploit the high-speed capabilities of SiGe HBTs, whereas the digital portion will utilize the high-density and low-power characteristics of MOSFETs. Pure SiGe HBT technologies also exist and are widely employed in the design of high-speed circuits. SiGe BiCMOS is a technology platform of its known and is distinct from CMOS. In terms of monolithic electro-photonic integration of a photodetector and TIA, however, SiGe BiCMOS and CMOS platforms are similar because they both have access to the implementation of a germanium photodetector. The successful demonstration of integrating germanium photodetectors into SiGe BiCMOS or CMOS optical receiver front-ends has long been reported [30], [31]. In [31], light is guided through silicon-on-insulate (SOI) waveguides and evanescently coupled to germanium, which is grown on top of the

waveguide. This is a popular approach because photon absorption occurs along the length of the waveguide, while the collection of charge carriers is perpendicular to this path. Therefore, the dependence between detector geometry and speed, responsivity is decoupled.

The integration of germanium photodetectors into a CMOS process has its shortcomings. The access to pure Ge growth in a single foundry is not always possible. Usually, post-processing will be done at a separate foundry, which is not economically feasible. Thermionic emission is another disadvantage of Ge photodetectors because it contributes to a relatively high dark current density [32]. Dark current contributes to detector shot noise, which will be discussed further in Chapter 4. Photodetectors built in silicon overcome the economic and logistical limitations of germanium detectors. The higher absorption coefficient silicon has at shorter wavelengths enables 850 nm, short-reach, optical interconnects suitable for datacenter and supercomputer applications. The combination of CMOS-compatible photodetectors alongside cheap and high-speed VCSELs presents a substantial case for short wavelength communications. Avalanche photodetectors (APDs) are often built in bulk CMOS technologies to boost responsivity. In bulk CMOS technologies, many carriers are generated outside the depletion region and slow diffusion current into the depletion region limits the speed of the detector.

2.6 Conclusions

The various semiconductor options that enable electro-photonic integration were discussed. Although photonic devices built in III-V platforms have superior performance to those in silicon, the dominance of silicon in digital electronics continues to push the development of high-performing photonics in silicon. However, the applications of monolithic electro-photonic integration in silicon will likely be limited to short wavelengths, whereas integration in III-V can theoretically span all communication wavelengths. Still the trade-off between responsivity and speed in pure silicon and germanium-based detectors limits the achievable speed to 25 Gb/s.

Chapter 3

3 Technology comparison

Being at the front of the receiver, the interface between the photo-detector and TIA, as well as the TIA itself, contributes significantly to the overall noise and bandwidth characteristics of the receiver. Semiconductor technology platforms such as InP, SiGe, and CMOS can facilitate the monolithic integration of electro-photonic circuits [33], [34], such as a PD and TIA. With miniaturization, the transition frequency of CMOS transistors (*e.g.*, 450 GHz for the 32 nm technology node) is becoming comparable to that of InP technology (520 GHz) [26 - 28] enabling higher speed circuits. It is therefore important to assess the current performance of these technologies as optical receiver front-ends, especially as signaling rates increase. As such, state-of-the-art optical receiver front-ends in each of these technologies are investigated in this chapter by using the common performance metrics of power consumption and sensitivity. For TIAs at comparable bit-rates and amplification, these metrics can be used to extract any competitive advantages between technologies. Receivers designed for 40 and 50 Gb/s are chosen for this study as they enable 100G, 400G, and 1600G data communication links (e.g., Ethernet) via advanced modulation schemes (e.g., PAM-4) and wavelength division multiplexing.

3.1 Receiver noise

The sensitivity of an optical receiver is directly related to the noise performance of all components of the optical link, especially the TIA. In this study, it is assumed that the TIA is the dominant

source of noise. TIA noise consists of thermal noise sources generated from transistors, resistors, etc. as well as current-dependent noise sources. However, an optical link will typically have other sources of noise including the relative intensity noise (RIN) σ_{RIN} from the laser, amplified spontaneous emission (ASE) noise σ_{ASE} if any fiber amplifiers are used, and shot noise σ_{PD} from the PD. In general, the total RMS noise value consists of a squared summation of these various noise sources, as described by equation 3.1.

$$\sigma_{rms} = \sqrt{\sigma_{TIA}^2 + \sigma_{PD}^2 + \sigma_{RIN}^2 + \sigma_{ASE}^2} \tag{3.1}$$

In equation 3.1, it can be seen why a dominant TIA noise source σ_{TIA} would approximately become the total RMS noise value σ_{rms}. Shot noise σ_{PD} from the PD is usually its dominant source of noise and is dependent on dark current I_d. For comparison purposes, a high-speed TIA built with modern technology might have an input-referred RMS noise current specification of 2 μA, or a mean-squared value σ^2_{TIA} of 4 x 10^{-12} A^2. Assuming there is no PD avalanche gain, the power spectral density (in A^2/Hz) of the shot noise of a photo-detector, as a function of dark current I_d and electron charge q is described below [18]:

$$\frac{d\sigma_{PD}^2}{df} = 2qI_d \tag{3.2}$$

Equation 3.2 would need to be integrated over a frequency range that depends on the filtering characteristics of the circuit blocks that follow the PD, such as the TIA and digital circuits. For a conservative approximation of how minor PD shot noise contribution is when it has low dark current, equation 3.2 will be integrated up to 50 GHz. Using a dark current value of 0.07 nA, which was experimentally obtained from a high-speed InGaAs PD and will be described further in Chapter 4, σ^2_{PD} equates to 1.1 x 10^{-18} A^2 – a millionth of the above calculated σ^2_{TIA} of 4 x 10^{-12} A^2. Thus, for high-speed applications, the TIA noise is almost always the dominant contributor of noise and the dark current dependent shot noise from the PD can often be ignored in receiver sensitivity formulations. The current power spectral density of resistive noise sources is proportional to *4kT/R* and since high-speed TIAs require low R, noise specifications have generally been rising with higher data rates.

3.2 Receiver sensitivity

Input signal sensitivity, power consumption, data rate, and footprint are important receiver front-end specifications. The input sensitivity is directly related to the noise performance of the TIA. As area is often compromised by the added footprint of optical sub-assembly (OSA), where the TIA and PD lie on separate dies and are wire-bonded together, reported state-of-the-art TIAs are compared in terms of their sensitivity and power consumption only at data rates of 40 and 50 Gb/s. Both metrics relate to the overall power consumed by an optical link – important for energy-efficient data communication products. The TIAs that were tested with a photo-detector report sensitivity as an average optical power, whereas the ones that were tested with a current source to emulate a PD, usually report an input referred RMS noise current. Assuming the TIA noise is dominant, the peak-to-peak current sensitivity of a receiver front-end ($i_{p\text{-}p}$) is directly related to the input referred RMS noise current (σ_{rms}) by taking the signal-to-noise ratio (SNR) required to achieve a specific bit error rate (BER) into account. Assuming Gaussian noise, the minimum SNR required to obtain a BER of 10^{-12} is 14 [38]. Based on photodetectors applicable to both 1310 and 1550 nm wavelengths, a photodetector responsivity (R) of 0.7 A/W is assumed. Then, the minimum detectable optical power (P_{sens}) at a given noise level is defined by equation 3.3 [39]. Note that this is also the average optical power, which is a quantity that can be tuned and measured with a power meter in a BER testing environment.

$$P_{sens} = 10\log_{10}\left(1000\frac{SNR\times\sigma_{rms}}{2R}\left(\frac{e+1}{e-1}\right)\right) \tag{3.3}$$

In the above expression, the extinction ratio (e) is the optical power ratio between the '1' and '0' power levels of an NRZ data stream. For simplicity, a high extinction was assumed for all TIAs in this study. The optical extinction ratio is limited by the transmitter of an optical link and is typically unreported in receiver literature. If the extinction ratio is greater than 9, for example, the sensitivity in equation 4.1 is compromised by less than 1 dB. If a low extinction ratio is used, the penalty is larger and the link budget suffers. However, assuming all receivers are to be utilized for the same optical interconnect standard implies that the extinction ratio should be constant. Hence the relative comparison of sensitivities between technology platforms will not change. Note that sensitivity can also be expressed in terms of an optical modulation amplitude (*OMA*), which

is the difference between the '1' and '0' optical power levels, as in equation 4.2. Theoretically, the BER depends only on the receiver noise, the responsivity, and the (*OMA*) that is present at the photo-detector.

$$P_{sens} = 10\log_{10}\left(1000\frac{OMA}{2}\left(\frac{e+1}{e-1}\right)\right) \tag{3.4}$$

3.3 Literature Survey

3.3.1 Transimpedance amplifiers

Figure 3.1 shows the optical power sensitivity with respect to the power consumption of reported TIAs using various technologies [40 – 51]. TIAs with equalizers are excluded due to their higher power consumption. Low-bandwidth CMOsS TIA approaches, such as in [52], are also excluded as similar designs are not reported in InP nor in SiGe BiCMOS. Of the twelve investigated designs, eleven of them took the parasitics of the front-end into consideration by either measuring receiver performance with a photodetector at the input, including an on-chip capacitor to emulate a monolithically integrated photodetector, or by adding parasitics in simulation. Only five of the twelve TIAs incorporated a photodetector. The assembly process introduces capacitive parasitics contributed by both the PD and TIA. Thus, the exclusion of these parasitics could result in an overestimation of the bandwidth reported by seven of the twelve designs explored in this study. In the next section and Chapter 5, the impact of excluding the full optical front-end will be elucidated. If equalization was used to mitigate bandwidth reduction, for example, a larger input capacitance requires an equalizer frequency response that accentuates the high frequency components of the signal [18]. As a result, more noise is integrated across the receiver chain and sensitivity is compromised.

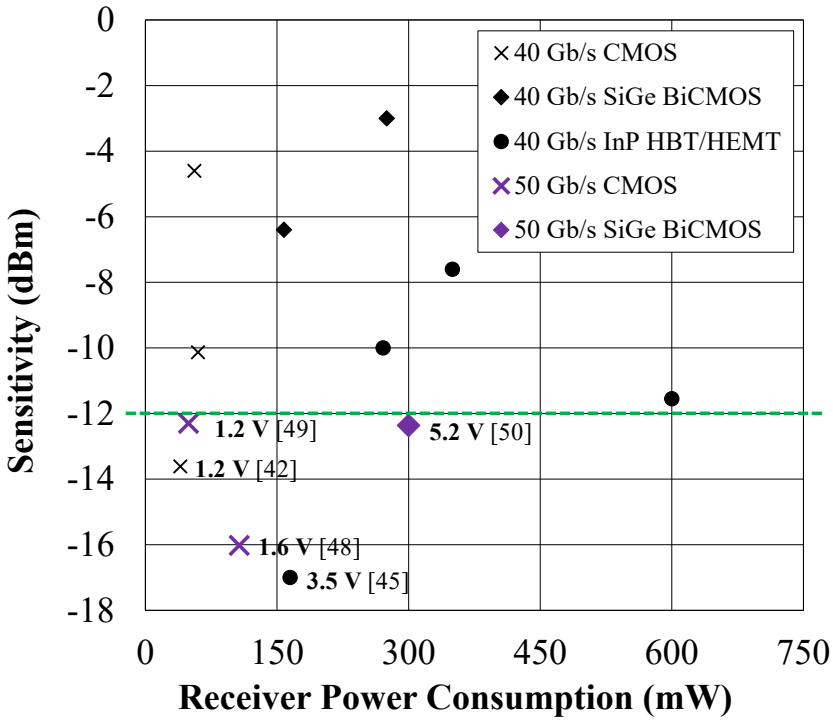

Figure 3.1 Sensitivity versus power consumption for 40 and 50 Gb/s TIAs reported in literature.

In Figure 3.1, all three technologies' receiver sensitivities are dispersed between -2 and -18 dBm, but the power consumed by CMOS designs are all within 150 mW. At sensitivities below -12 dBm, CMOS designs [42], [49] consume approximately a third and a fifth of the power of InP bipolar [45] and SiGe BiCMOS [50] designs, respectively. In fact, the bipolar receiver [45] and BiCMOS receiver [50] utilize supply voltages that are 2.7 and 4.3 times greater than the CMOS receivers [42], [49], which is roughly the same order of magnitude as the power consumption difference. The TIA in [45] uses a supply of 3.5 V, which is more than twice as large as CMOS design's supply in [48], but consumes only 1.5 times the power. This suggests that bipolar technologies' power consumption may be ultimately limited by higher supply voltages. In terms of current consumption, these results are consistent with theory: bipolar transistors require less current than MOSFETs to achieve a given transconductance [2]. However, it is important to note that only collector current in bipolar transistors and drain current in MOSFETs are considered in this assessment. Base current in bipolar devices and gate leakage current in MOSFETs also contribute to power consumption. Thus, while a CMOS TIA can be inferior in terms of gain, when

compared to a SiGe TIA [2], their lower supply voltage and ability to integrate clock and data recovery (CDR) circuits on-chip, make them favorable from an energy standpoint.

3.3.2 Area considerations

Although CMOS-based TIAs can theoretically be made more compact than bipolar technologies, due to constant transistor scaling, an external PD fabricated using a III-V process is typically required to achieve data rates above 25 Gb/s. As described in chapter 2, the photonic properties of silicon at 1310 and 1550 nm do not permit efficient light detection, further necessitating the requirement of an external PD.

The assembly of photonic and electronic components via wire-bonding makes the overall receiver less compact. The bond pad of the PDs that will be characterized in chapter 4, for example, are 75 µm x 75 µm with a pitch of 100 µm. Furthermore, commercial PDs typically have three bond pads, with the centre being for the signal and the two outer bond pads for the power supply or ground. For wire-bonding purposes, a minimum bond pad size of 60 µm x 60 µm is required on the IC side (i.e., the TIA). Following the Canadian Microelectronic Corporation's (CMC) recommendations for wire-bonding, a PD-to-TIA distance of at least 150 - 500 µm is realizable. With more optimistic bond pad dimensions of 60 µm x 60 µm for the PD's 3 bond pads and a minimum PD-to-TIA distance of 150 µm, an area of at least 0.07 mm^2 is added by optical sub-assembly. This is shown in Figure 3.2.

For a longer wire-bond length and a PD-TIA PD-to-TIA distance of 500 µm, 0.16 mm^2 is added to the optical receiver's overall footprint. spacing of Note that this ignores the area from the pad to the edges of the chip. By contrast, a single-stage TIA designed in chapter 5 consumes 5.4×10^{-6} mm^2.

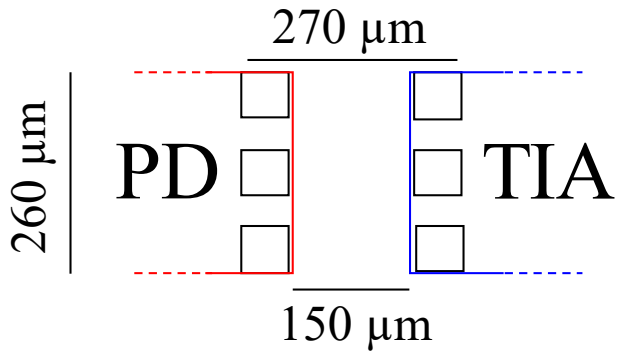

Figure 3.2 Minimum area of 0.07 mm^2 added by PD-TIA optical sub-assembly via wire-bonding

3.3.3 CMOS photodetectors

Although CMOS-based TIAs can theoretically be made more compact than bipolar technologies, due to constant transistor scaling, an external PD fabricated using a III-V process is typically required to achieve data rates above 25 Gb/s. This renders the overall design less compact. While the area of the photodetector itself can be small (e.g., a radius of 5 µm in [53]), the area added by the bond pads (e.g., 60 µm x 60 µm each, with a pitch of 100 µm between bond pads) of both the PD and TIA chips, and the wire bonding length (in the order of 150-500 µm) potentially add 0.16 mm^2 to the total area of the optical receiver front-end. It is worth noting that while smaller PDs, such as the one reported in [53], achieve a bandwidth as high as 62 GHz by significant reduction of the junction capacitance, this simultaneously reduces their responsivity. For compactness, flip-chip methods can be used to interface the PD to the TIA, resulting in reduced parasitic inductance at the input. However, the total input capacitance of the TIA increases [54]. The impact of flip-chip bonding will be explored in more detail in the next section.

For a CMOS TIA to be monolithically integrated with a PD, Ge absorber layers should be incorporated into the CMOS process. This is not a straightforward or trivial change and is therefore currently limited to legacy fabs running relatively large node size processes. The PD reported in [55] is integrated in a 180 nm CMOS node and achieves a bandwidth of 8.5 GHz, the PD in [56] is integrated in a 130 nm CMOS node and achieves a bandwidth of 18 GHz, and the monolithically integrated receiver implemented in a 90 nm CMOS node reported in [57] achieves a data-rate of 25 Gb/s. Hence, the fastest reported monolithically integrated CMOS receiver achieves only 25

Gb/s. The main limitation for having a CMOS process that could be adapted for optimized photonic and electronic performance is the access to pure Ge growth [29]. The economic feasibility of this process amendment has yet to be realized, especially at 32 and 28 nm node sizes. InP and SiGe technologies provide less complex monolithic approaches.

3.4 Device scaling and f_T

The speed of an integrated circuit technology is usually defined by its transition frequency f_T. For MOSFET devices using a CMOS process, it is the frequency at which a small-signal gate current i_g equals the small-signal drain current i_d. For HBT devices using a SiGe BiCMOS or InP HBT process, it is the frequency at which a small-signal base current i_b equals the small-signal collector current i_c. In other words, the f_T is the cut-off frequency at which the current gain of a transistor equals unity [24].

The f_T is proportional to the transconductance g_m of a transistor, which was defined in chapter 2 as the small-signal voltage-to-current gain. It is inversely proportional to the inherent capacitive parasitics and any delays caused by the transit time of carriers in the transistor. In the CMOS process, this is the gate-source C_{gs} and gate-drain C_{gd} capacitance. In HBT technologies, this includes the emitter and collector junction capacitance, C_{je} and C_{jc}, and the carrier transit time τ of electrons or holes in the base region [17]. These junction capacitances result from the diode nature of the npn or pnp interface and will be defined in chapter 4 as part of the characterization process of photodetectors, which are also diode-based devices. In an npn transistor, the f_T is inversely proportional to the transit time of electrons in the p-type base region. As discussed in chapter 2, the higher mobility of electrons enables faster transit times than holes, which is why npn transistor are preferred over pnp transistors for high-speed applications. The f_T for both MOSFETs and bipolar devices is summarized in equations 3.5 and 3.6, respectively [17].

$$f_T = \frac{1}{2\pi} \cdot \frac{g_m}{C_{gd} + C_{gs}} \tag{3.5}$$

$$f_T = \frac{1}{2\pi} \cdot \frac{1}{\tau + (C_{je} + C_{jc})/g_m} \tag{3.6}$$

As described by equations 3.5 and 3.6, the scaling down of transistor size benefits the high-speed performance of both HBT and CMOS technologies due to reduced device capacitance. This is in addition to the benefits of higher integration density, which lowers the cost of a die. As such, it is a trend for both bipolar and CMOS processes to employ fabrication innovations that reduce device dimensions via improved lithography.

It is well known that the channel length of MOSFETs has been halved every two to three years or so [21]. In BJTs, technology scaling typically reduces the thickness of the base and emitter regions and the length of the emitter. A thin base decreases τ in equation 3.6 and smaller emitter lengths reduce C_{je} [19]. In HBT devices, τ is further reduced due to the stronger electric field created through the reshaping of energy bands due to smaller-bandgap base materials such as InGaAs and SiGe. This is the primary principle by which a heterojunction bipolar transistor boasts higher clock rates than a conventional bipolar junction transistor.

Regardless of aggressive technology scaling in CMOS devices, the g_m of a MOSFET is also proportional to its width-to-length W/L ratio as described in equation 3.7. Therefore, the size W of transistors cannot be made too small for an optimal transconductance. This is especially important in high-speed analog circuits like TIAs, where transconductance can help with f_T and the open-loop gain A of some of the amplifier topologies discussed in chapter 2. In equation 3.7, V_{GS} is the gate-source voltage, V_t is the threshold voltage, μ_n is the carrier mobility, and C_{ox} is the capacitance of the thin oxide or insulating layer between the gate and body of the transistor [2].

$$g_m = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_t) \tag{3.7}$$

This may explain why, despite consistent technology scaling in CMOS, the to-scale images in Figure 3.3 of a 50 Gb/s InP HBT TIA (right) built in 2003 [46] and a 65 nm CMOS TIA (left) reported in 2014 [49] approximately consume the same area. In Figure 3.3, both TIAs include an initial transimpedance stage followed by differential post-amplifier circuits, and a 50 Ω output buffer that impedance matches to high-speed testing equipment or 50 Ω transmission lines on a board. Thus, both circuits are primarily analog.

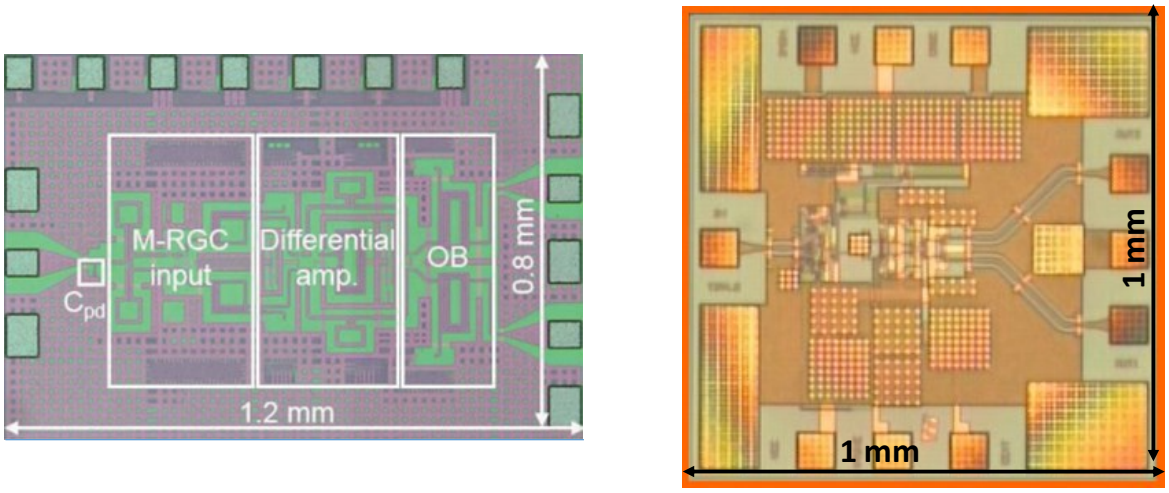

Figure 3.3 Chip micrographs of a 65 nm CMOS TIA (left) [49] and an InP HBT TIA (right) [46].

Unlike MOSFETs, the g_m of bipolar devices has less of a size dependence. This is exemplified by equation 3.8 below, where I_C is the collector current, k is the Boltzmann constant, T is temperature, and q is the charge of an electron. Together, kT/q is equal to the thermal voltage V_T, which is about 26 mV at 300 K [24].

$$g_m = q\frac{I_C}{kT} = \frac{I_C}{V_T} \tag{3.7}$$

Like MOSFETs, BJTs also trade off important characteristics when the devices are made too small. For example, smaller devices will exhibit higher collector current density (A/m^2) for a given collector current I_C. The collector current density J_c is taken as the collector current divided by the emitter area, which is the product of its width and length. For every BJT or HBT, there is an optimal collector current density that maximizes f_T and minimizes noise [58]. While the collector current is determined by the requirements of the circuit (i.e., g_m needs to be maximized), the electronic designer will want to size their devices near these optimal J_c values. As the thickness and length of the bipolar emitter is decided by the fabrication technology, the width of the transistor is left as a design variable. The aforementioned trends and device capacitances for MOSFET and bipolar devices is summarized in Figure 3.4. The sizing trends outlined in green are favourable for

footprint reduction whereas those in red typically increase the footprint of the IC. To reiterate, scaling up the width of the transistor by relative to the channel length (which scales down due to technology) in CMOS is largely beneficial in high-speed design because it increases the transconductance g_m. Of course, there is an upper limit, beyond which device capacitance increases and reduces the transition frequency f_T. In bipolar devices, the width is not always increased but, instead, sized to obtain an optimal collector density that minimizes noise.

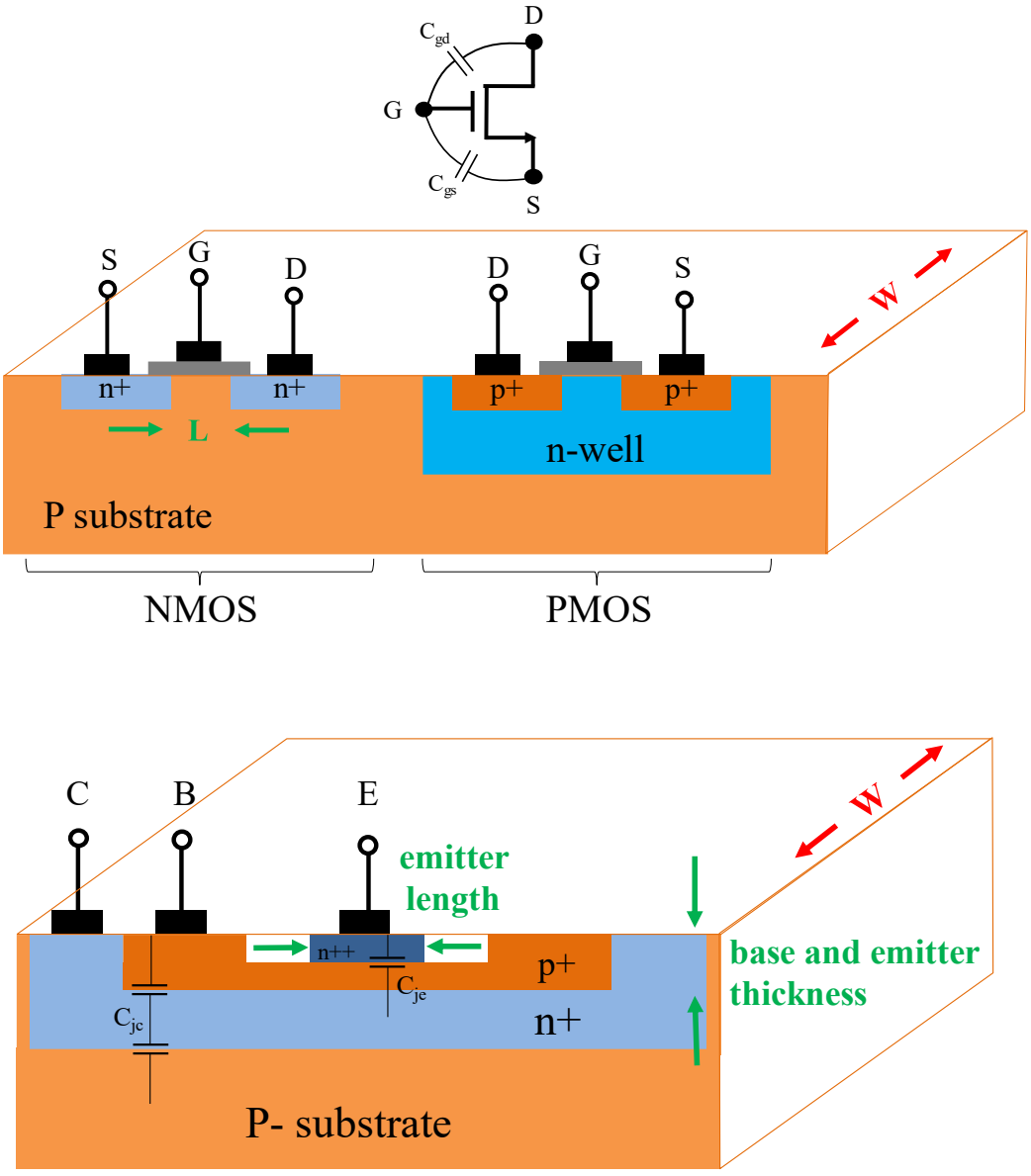

Figure 3.4 Scaling trends, device capacitances, and sizing of silicon MOSFETs (top) and InP or SiGe HBTs (bottom) for optimum high-speed performance.

Chapter 4

4 Experimental photodetector characterization

In this chapter, a III-V photo-detector is experimentally characterized at single-mode, communication wavelengths of 850 nm, 1310 nm, and 1550 nm. It is then wire-bonded to a TIA. The results and the knowledge gained from this characterization process will directly be used in the quantitative discussion of receiver front-end sub-assembly versus monolithic integration.

The photodetector characterization project was part of an industrial collaboration with INTENGENT Inc. – a III-V semiconductor start-up based in Ottawa, Ontario. Three different photodetector chips were fabricated at Global Communications Semiconductors (GCS): one was a 25 Gb/s commercial detector developed by the foundry itself and the other two were designed by the industrial collaborator. All detectors were grown using GCS's semi-insulating (SI) InP substrates and differ only by the design of the vertically-grown epitaxial layers. The two different designs proposed by the collaborator was an initial attempt at understanding and overcoming (in future designs) the common responsivity-bandwidth trade-offs of photodetectors. Although the experimental results of the characterization will be discussed in-depth, the specifics of the epitaxial layers of all three designs are proprietary and will not be disclosed. The goal of the industry collaborator was to devise a monolithic receiver front-end using GCS's III-V process. This would be a competing solution to some of the silicon photonics and SiGe electro-photonic circuits discussed in Chapter 2.

A cross-section of the commercial detector, hereinafter referred to as C3 devices, and micrograph of the top of the device is shown in Figure 4.1. The device operates by vertically coupling light through the circular aperture shown in the micrograph. By collecting experimental information, the goal is to build an electrical model of the PD-TIA interface.

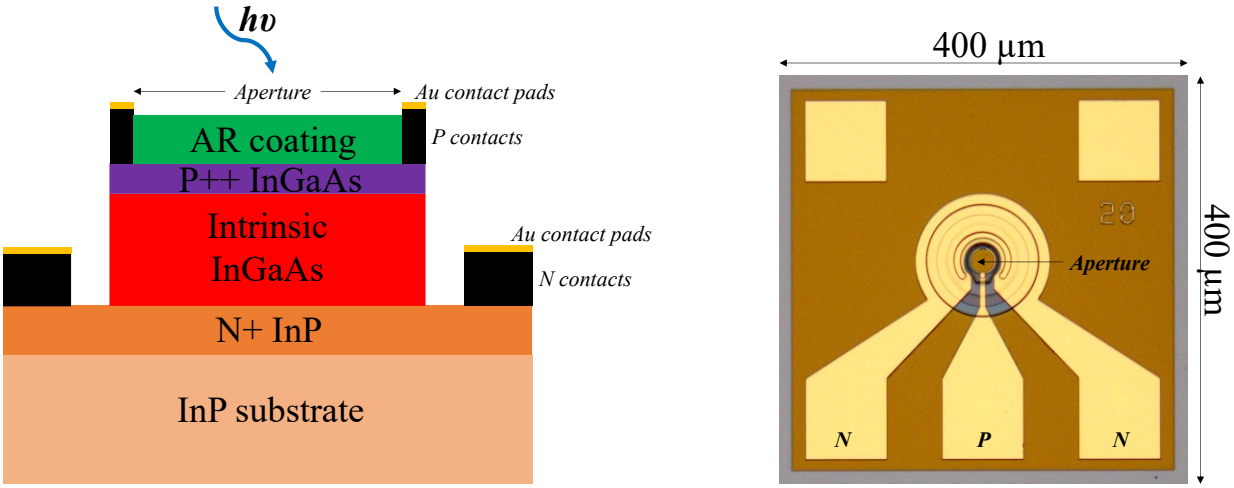

Figure 4.1 Cross-section (left) and top-view micrograph (right) of the commercial photodetectors tested.

4.1 Capacitance-voltage

When an increasing reverse bias voltage is applied across a diode, its depletion region widens. The junction capacitance, then, is inversely proportional to the depletion region width – akin to a parallel plate capacitor. Thus, the junction capacitance of a photodetector can be reduced by increasing the bias voltage, as described by (4.1) [10].

$$C_j = \frac{A}{(V_i + |V_r|)^{\frac{1}{2}}} \left(\frac{q\varepsilon(N_a N_d)}{2(N_a + N_d)} \right)^{\frac{1}{2}} \tag{4.1}$$

where V_i is the built-in potential, V_r is the applied reverse bias, A is the area of the junction, q is the electron charge, ε is the permittivity of the semiconductor, N_a is the acceptor (p or p+) doping concentration, and N_d is the donor (n or n+) doping concentration. Due to the junction

capacitance's inverse square root dependence on the applied bias V_r, the capacitance reduction per voltage, which can be quantified by taking the derivative of equation 4.1, has diminishing returns. Specifically, it is proportional to $V_{total}^{-1.5}$, where V_{total} is the sum of V_i and V_r. Thus, an RC-limited detector's bandwidth cannot be steadily increased by simply increasing V_r.

The exact nature of this relationship between capacitance and voltage depends on the detector design and configuration and can be determined by performing capacitance-voltage (CV) measurements, where the total capacitance across the photodetector is measured as a function of the applied voltage. In the case of a vertical p-i-n structure that is more representative of these detector designs, for example, the capacitance is inversely proportional to the width of the intrinsic region, which is theoretically fixed with little bias-dependence. Contrasting the junction capacitance, the parasitic pad capacitance is bias and intrinsic/depletion area-independent. It depends on the geometry of the pads and semiconductor-electrical contacts, which was the same for all three detector designs in this case.

The total capacitance can be de-embedded into the junction and pad capacitance. Fortunately, the industrial collaborator's designs featured aperture diameters of 16 µm and 20 µm, allowing one to empirically determine the relationship between total capacitance and aperture size. Linear regression can then be used to determine the parasitic pad capacitance, which is the capacitance that remains in the hypothetical limit of reducing the aperture diameter to 0 µm. This process is described by (4.2), where the second term in the summation is the junction capacitance. It has a linear dependence on the aperture area A through the slope α, which is composed of the permittivity and thickness of the intrinsic region.

$$C_{total} = C_p + \alpha A \tag{4.2}$$

$$C_{total} = C_p + C_j \tag{4.3}$$

Through this analysis, it was found that parasitic pad capacitance derived from testing multiple devices of the three different designs was within the range of 45 – 50 fF. As expected, the difference in junction capacitance between the 20 and 16 µm diameter photodetectors is directly proportional to their areas: $20^2/16^2$. Using the regression method, the junction capacitance of the commercial C3 detectors is 55 fF.

The approximate diameter of the core of single-mode fibers at short and long communication wavelengths are 5 μm and 10 μm, respectively. Therefore, a properly-aligned single-mode fiber should be able to couple most of the light into the 16 and 20 um photodetector apertures discussed in this chapter. Clearly, these detectors are better suited for single-mode fibers and hence, this work was done using single-mode systems only. Multi-mode fibers, which are usually optimized for 850 nm communications, have diameters that range from 50 to 100 μm. Therefore, detectors designed for multi-mode fiber communications would have to be made larger than single-mode detectors, which imposes a tighter design constraint on the speed-responsivity trade-off.

4.2 Continuous-wave measurements

In continuous-wave (CW) measurements, the light source from a laser is directly coupled to the photodetector without data. In doing so, the responsivity of the photodetector is attained by monitoring the current drawn by the photodetector and dividing it by the power that is coupled into the detector. A schematic of the test-bed used for these experiments in described in Figure 4.2. The responsivity of the C3 detectors is measured with respect to the applied reverse bias voltage as shown in Figure 4.3. A variable optical attenuator (VOA) was placed in the set-up to provide extra control over the optical power going into the PD.

Using the same set-up described in Figure 4.2, the laser source was also turned off to record the dark current of the detector under reverse bias. The source was left on to generate the forward-bias current-voltage (IV) curves in Figure 4.4 to clearly identify the boundary at which the device crosses from the photo-detecting to forward current region. Two different devices of the same design were plotted in Figure 4.4 to evaluate fabrication variations, which appears to be minimal. Both devices can theoretically be operated as photodetector until a forward bias of approximately 450 mV. Low dark currents and expected exponential behaviour of forward IV curves helps identify which of the many PD dies are functioning well and can thus be used for further CW and high-speed testing. The best-case dark current and average responsivity (over 10 V of reverse bias) of a properly-functioning C3 PD is recorded in Table 4.1. As described in Chapter 3, the magnitude of this dark current is too low for it to significantly impact the performance of a high-speed TIA.

The lower responsivity of the detector at short wavelengths is attributed partly to the fact that the anti-reflective coating is optimized for 1310 and 1550 nm wavelengths and not for shorter wavelengths. From a system point-of-view, the process of simply switching an optical link's operation from long to short wavelengths (by changing the laser source alone) in a data center can cost almost 3 dB in the link budget if the same detector were to be used. In reality, drastic wavelength changes such as from O and C bands to short, is not trivial and requires an overhaul of almost all electro-optic hardware due to the non-broadband nature of most devices. As will be shown in the high-speed measurement section, these detectors are still capable of detecting data above 25 Gb/s at all three wavelengths, regardless of the responsivity reduction at short wavelengths.

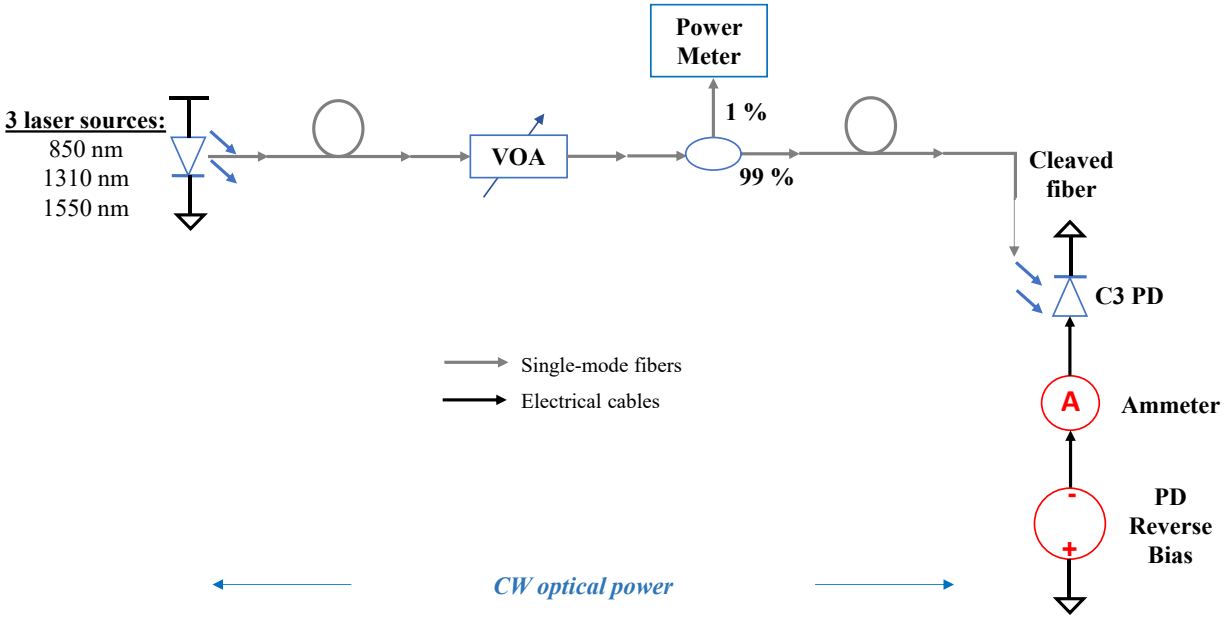

Figure 4.2 Schematic of experimental set-up for continuous-wave PD measurements.

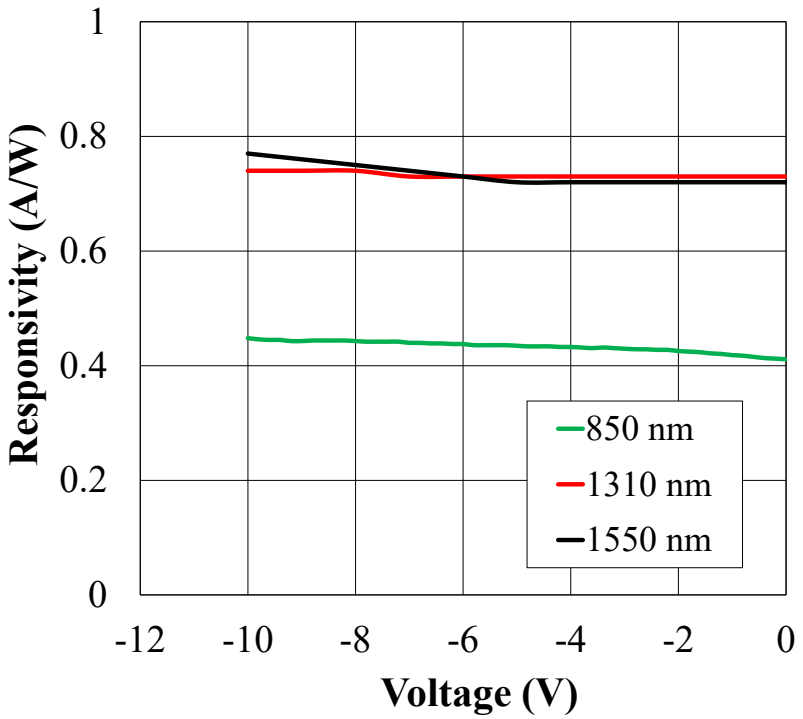

Figure 4.3 Responsivity versus voltage of the III-V C3 photodetectors at various wavelengths.

Table 4.1 Responsivity and dark current of III-V photodetector.

Wavelength (nm)	Average Responsivity (A/W)	Dark Current (nA)
850	0.43	
1310	0.73	0.07
1550	0.73	

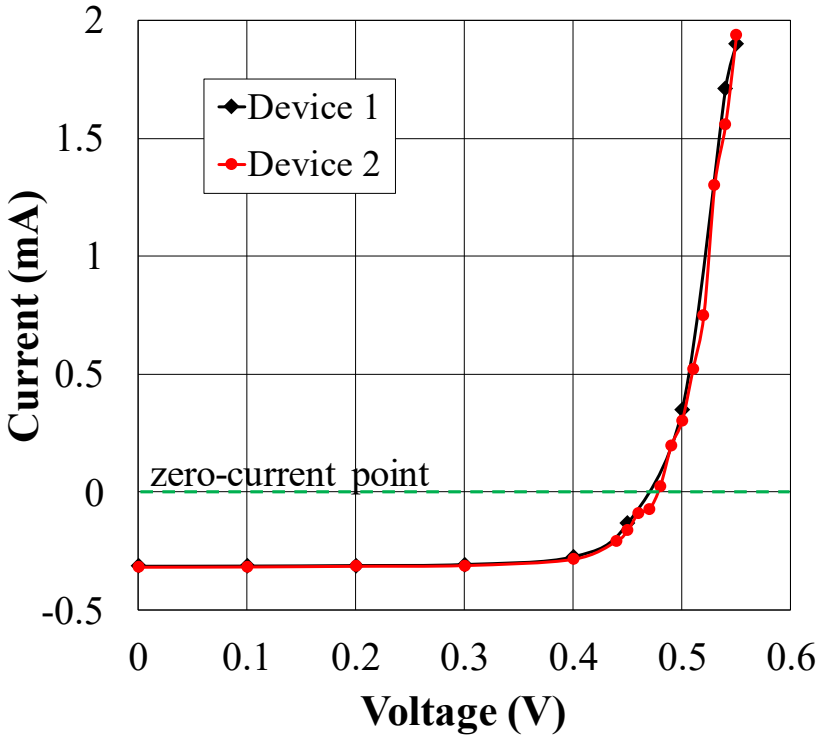

Figure 4.4 Forward I-V curves for two different (red and black) photodetector dies of the same design: C3. The zero-current point, where the devices change from a detector to a forward-biased diode, is outlined (dashed horizontal green line).

4.3 High-speed measurements

4.3.1 Small-signal

The small-signal O/E frequency response of the photodetectors was determined by taking S_{21} parameters at various bias voltages. The S_{21} parameter quantifies the power gain (ratio of output power) of the system, which, in this case, is the photodetector. Since this is an O/E device, the input power is optical and the output power is electrical. Thus, these measurements were enabled by using a 50 GHz vector network analyzer (VNA) in conjunction with a light-wave component analyzer (LCA) that functioned at both 1310 nm and 1550 nm. The vector network analyzer sweeps sinusoidal frequencies from DC to 50 GHz and feeds these electrical signals into the LCA, which has a built-in E/O modulator for electro-optic conversion. The light coming out of the LCA is then vertically-coupled to the PD. As long as there is sufficient light coupling into the PD to generate an electrical current that can be sensed by the VNA, rigorous fiber-to-PD alignment for minimal

loss is not essential. This is because the O/E frequency response (in dB) is normalized to the DC value. What matters is how the response decreases in magnitude as the frequency is swept. Specifically, the 3 dB point in the S_{21} curves is taken at various reverse bias voltages and plotted in Figure 4.5.

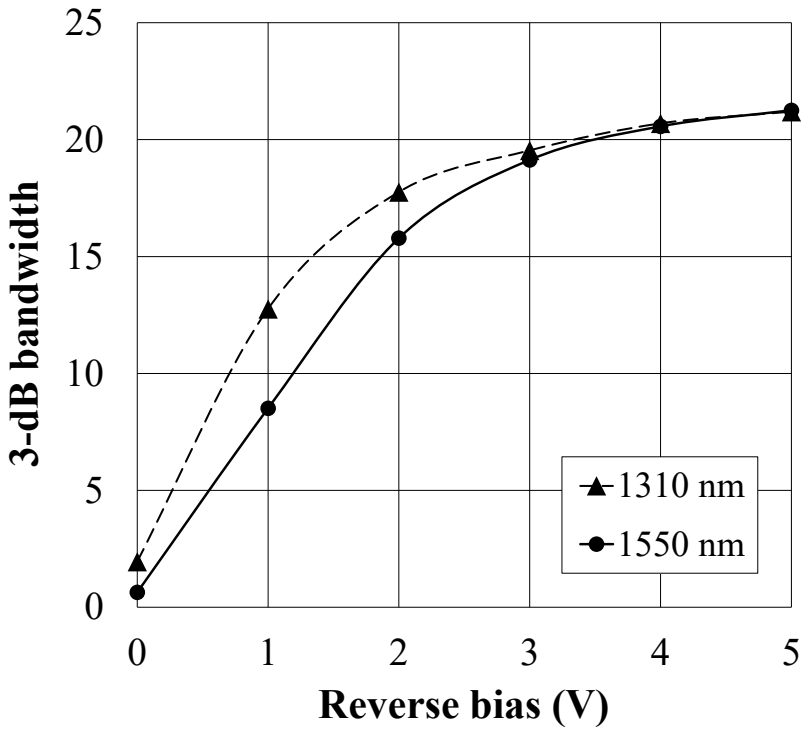

Figure 4.5 Bandwidth versus reverse bias voltage for C3 photodetectors. Results extracted via O/E S_{21} parameters.

The wavelength dependence of the C3 detectors stems from the fact that they are transit-time limited devices. Since every wavelength has its own photon absorption profile in the intrinsic InGaAs region, the relative location where the highest density of photons is absorbed and drift into the non-depletion region, will vary from wavelength to wavelength. Furthermore, increasing the reverse bias voltage exposes carriers such as holes and electrons to higher electric fields, allowing them to reach higher drift velocities (correlating to lower transit time) until they reach their saturation velocity. This explains why the curves in Figure 4.5 appear to be saturating as the bias voltage is increased. The transit-time dependent nature of these photodetectors was confirmed by

the industrial collaborator, INTENGENT Inc., by modeling the drift velocity of holes, which is the limiting carrier, as a function of reverse bias voltage. The model revealed the same trajectory and saturation characteristics as the bandwidth versus voltage curves in Figure 4.5.

Figure 4.5 yields valuable biasing information for the electro-photonic designer. As described in Chapter 2, the p-junction of the photodetector is connected to the input of the TIA. Therefore, the TIA or its biasing network sets the anode of the PD, while the cathode voltage is taken from either its own supply or the supply network of the TIA. Clearly, the PDs characterized in this chapter would be optimally biased at 3 or 5 V. However, many high-speed electronic devices and their processes can only withstand 3.3 V or less, making the design of monolithic receivers and/or electro-photonic sub-assembly biasing networks less straightforward. The input of the TIA is usually turning on a gate of a CMOS transistor or forward-biasing the p-n junction or a BJT. The 1.2 V, 65 nm CMOS process used in the design of the TIA in chapter 5 requires the gate voltage of the input transistor to be greater than 300 mV, which would limit a PD's bias voltage to 0.9 V if it were to be supplied from the TIA's supply network. Biasing a PD with the bandwidth-voltage characteristics as Figure 4.5 with a 1.2 V CMOS process as opposed to a 3 V III-V process could sacrifice its bandwidth by almost a half. Luckily, the optical sub-assembly process introduces a wire-bond inductance that helps extend the bandwidth of the overall receiver and isolates the PD capacitance from the TIA's input capacitance. On-chip inductors and other bandwidth boosting techniques are often employed on the TIA as well to further help with the bandwidth.

4.3.2 Eye diagrams and BER curves

In these sets of high-speed measurements, pseudorandom binary sequence (PRBS) non-return-to-zero (NRZ) data is actually generated via a bit pattern generator and detected by the photodetectors with the help of electro-optic (E/O) modulators. The resulting electrical current is then monitored by a digital component analyzer (DCA) that generates eye diagrams of the data. Additionally, the data from the PD is also monitored by a bit error rate tester (BERT) in order to assess the bare sensitivity of the detectors without a TIA. Measurements were performed at 850 and 1550 nm. A schematic of the test-bed is shown in Figure 4.6.

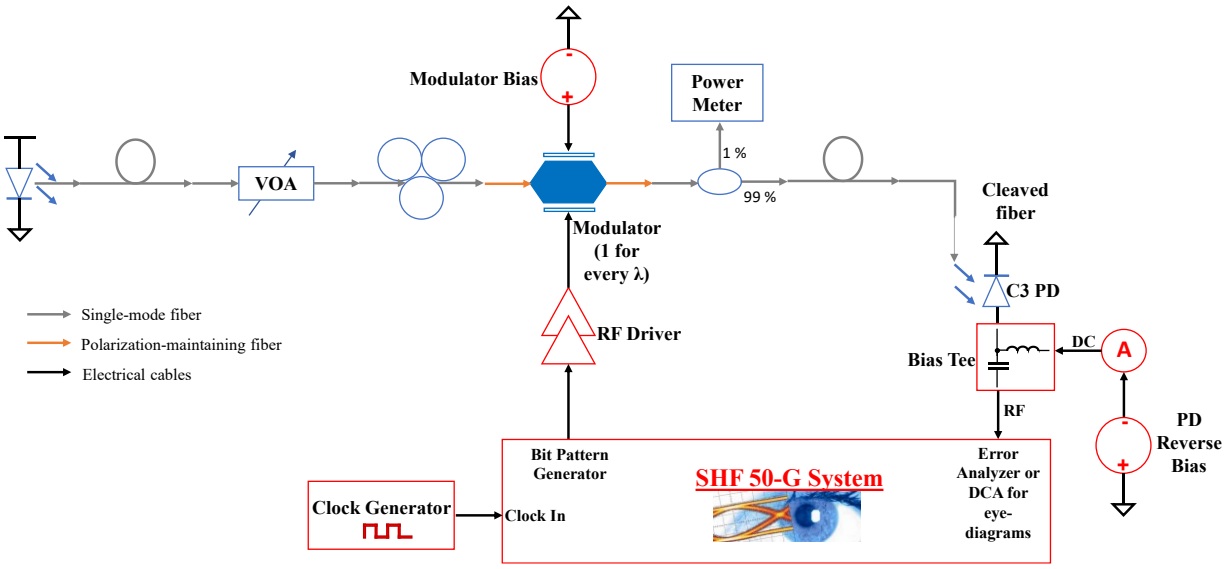

Figure 4.6 Schematic of experimental set-up for high-speed photodetector measurements.

The eye diagrams at both 850 and 1550 nm reinforce the bandwidth versus voltage trend observed in Figure 4.5. For example, at 28 Gb/s, the SNR of the NRZ data coming out of the photodetector at 1550 nm improves from 3.5 to 4.4 when the reverse bias is increased from 2 V to 5 V. This is shown in Figure 4.7 a) and b).

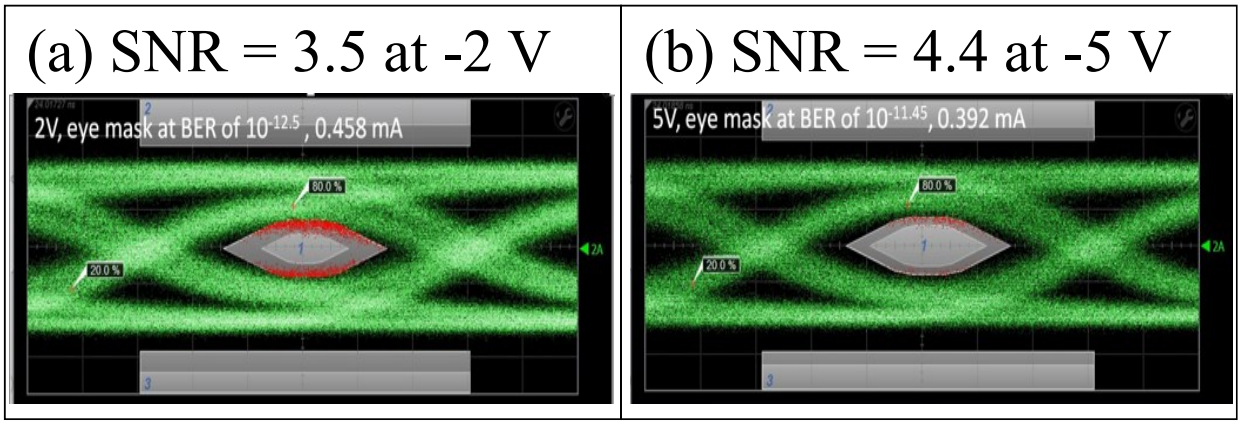

Figure 4.7 Eye diagrams extracted from photodetector at 1550 nm for 28 Gb/s PRBS NRZ data stream. a) Photodetector is biased at -2 V. b) Photodetector shows slightly improved response at -5 V bias.

The improvements in the eye quality when biasing the PDs at higher reverse bias voltages is also observed at short wavelengths, as demonstrated in Figure 4.8 below, where the bias voltage is increased from 0 to -5 V from the left-most to right-most image.

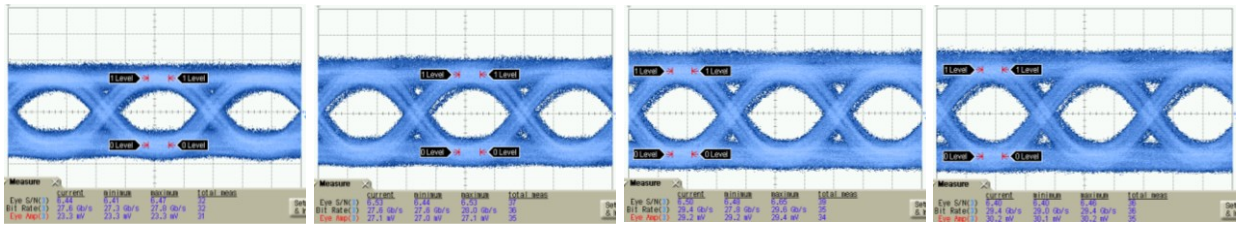

Figure 4.8 Eye diagrams extracted from photodetector at 850 nm for 28 Gb/s PRBS NRZ data stream. Detector bias is increased from left to right as follows: 0 V, -1 V, -3 V, and -5 V.

Table 4.2 summarizes the SNR results. The improvements in SNR versus PD bias are more marginal in this case and a peak is observed at -3V. At -5V, not only have the carrier drift velocities saturated, but it appears that there may even be a slight dip in SNR. This observation is further supplemented by the BER curves in Figure 4.9, where a bias of -3 V shows optimal performance.

Table 4.2 28Gb/s SNR versus voltage for photodetectors at 850 nm

Applied voltage (V)	Eye diagram SNR
0	6.44
-1	6.49
-3	6.57
-5	6.43

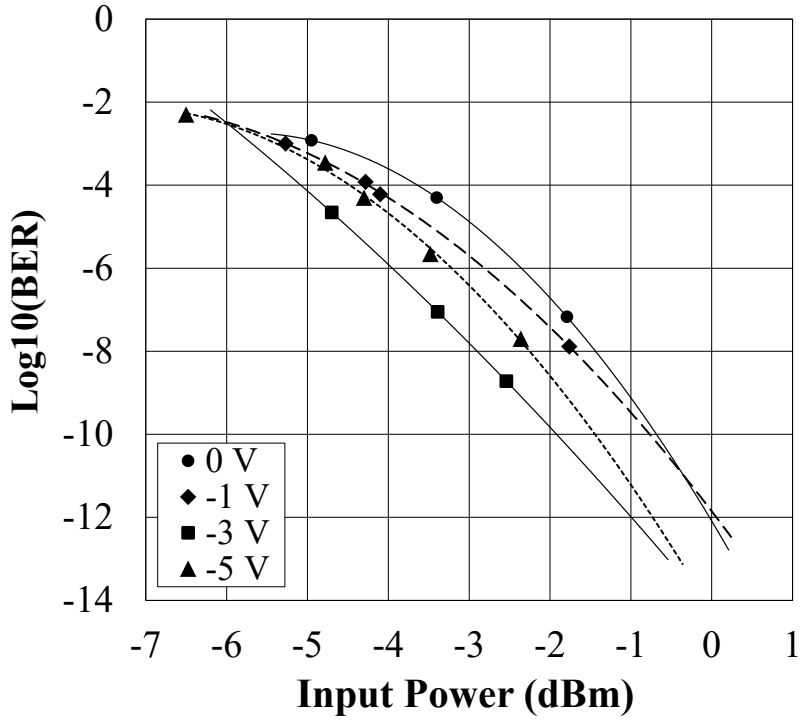

Figure 4.9 Measured BER as a function of reverse bias voltage for C3 photodetectors at 25 Gb/s, 850 nm.

4.4 Optical receiver front-end packaging

4.4.1 Assembly process

The process of packaging the characterized photodetectors with a commercial TIA was undertaken. In order to try and make the assembly as compact as possible, PD and TIA dies were directly bonded to a printed circuit board (PCB). The board was custom-designed with a wire-bond pad arrangement that all the power and ground required by the dies. A dedicated package was not used. The output of the PD was directly bonded to the input of the TIA. The TIA had a differential output that was probed in the lab by a high-speed ground-signal-signal-ground (GSSG) probe. A picture of the PCB, with a zoomed-in micrograph of the PD and TIA dies, is provided in Figure 4.10.

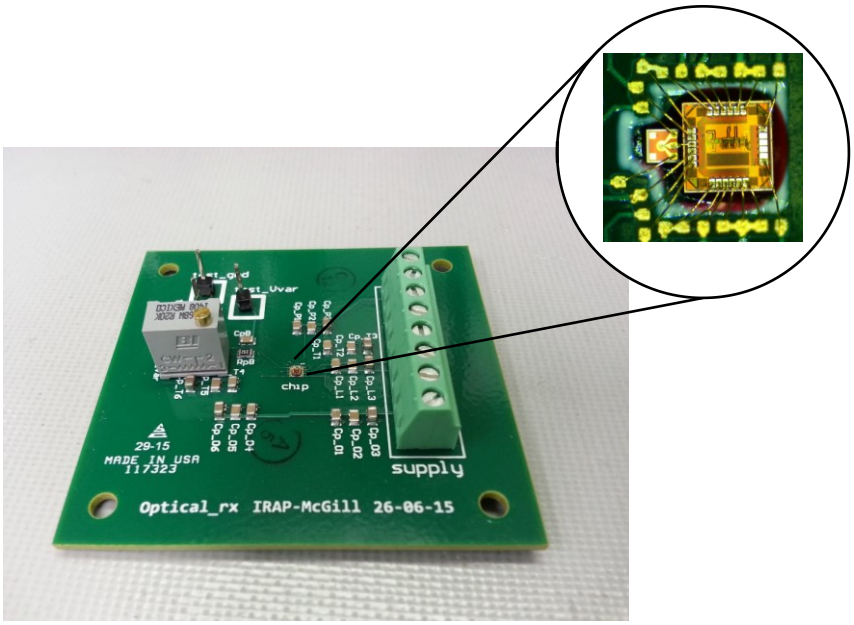

Figure 4.10 Image of complete optical receiver front-end package.

In going through this process, the constraints of wire-bonding and the approximate length of the wire-bond between the PD and TIA were determined. This helped set the context for the packaging analysis work that follows in the remainder of this thesis. A zoomed-in micro-graph of the PD-TIA interface is shown in Figure 4.11. By approximating the wire-bond as a semi-circular arc, the length between the PD and TIA is about 350 µm, corresponding to approximately 350 pH of inductance; as estimated from the general rule of thumb of 1 nH/mm of wire-bond.

Reducing the wire-bond length as much as possible is not a straight-forward task. These are few of the limitations: bond-pad to bond-pad distance, distance between the bond-pads and edge of the chip. The larger the height difference between the chips; the further they must be placed from each other to avoid a sharp wire-bond arc.

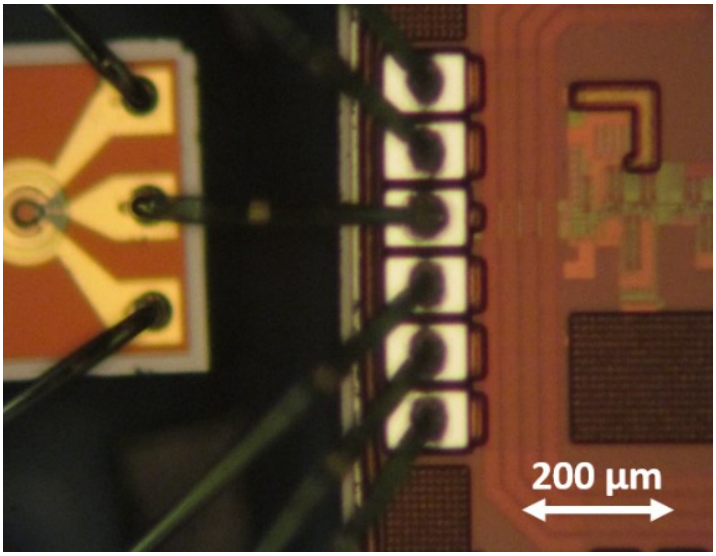

Figure 4.11 Micrograph of a PD wire-bonded to a TIA.

Board design and packaging adds extra cost and time to the optical sub-assembly approach. Monolithic integration may have larger upfront costs but could possibly be comparable to sub-assembly once the expense and time associated with packaging is considered. With designing TIAs in CMOS, the use of a monolithically-integrated versus a discrete photodetector may even result in selecting a 65 nm node (cheaper and lower f_T but transistor speed requirement is less due to a smaller input pole) versus a more expensive 45 nm node. This is because the reduction in bandwidth from the large input pole introduced by packaging a photo-detector and TIA, is compensated for by using a node where transistors are capable of operating at higher speeds.

4.4.2 Results

A total of five receivers were assembled on 5 PCBs. The receivers were tested with a 25 Gb/s NRZ data using a similar experimental set-up described in the high-speed photodetector characterization section. The output was monitored on a DCA in order to potentially collect eye diagrams before proceeding with BER testing. Even with the use of bypass capacitors placed close to the receiver for every power supply trace, a lot of noise was present at the output. The magnitude of the noise grew as the ground-signal-signal-ground (GSSG) probes were dug more deeply into the differential output pads of the TIA. Specifically, the output of the 50 Ω driver was probed. In fact,

the output changed in response to the vibrations of the room and the exact position of the probe, suggesting there is likely an aluminium oxide layer on top of the pads that needs to be penetrated for a proper contact.

Two major issues (each board with a different issue) were found from the tested receivers:

1) Noise that did not change with any input stimulus.

2) Hardly resolvable low-magnitude waveform covered in much noise and jitter, as shown in Figure 4.12. The eye changes slightly with input stimulus, such as increasing the optical power present at the photo-detector or the data rate. The snapshot in Figure 4.12 yields an SNR of 3.62 but this value deteriorates with time due to eye closure.

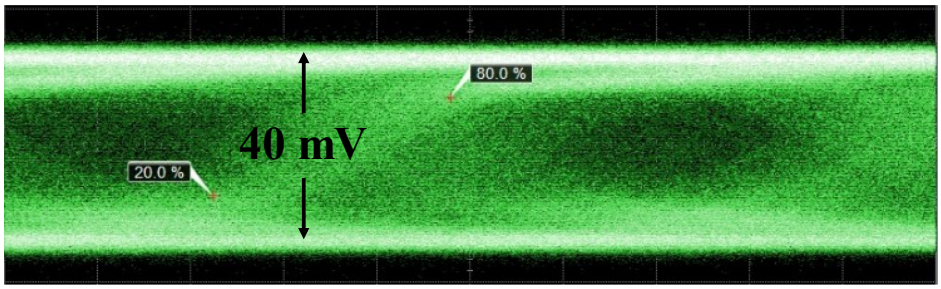

Figure 4.12 A 25 Gb/s noisy eye diagram with a lot of jitter for an assembled receiver front-end.

4.4.3 Failure analysis & prevention

The experimental and assembly conditions were considered more in-depth after the testing process. In order to ensure a more robust receiver package with potential for reportable results, the followings measures should be considered:

1) A noisy output that does not change with any stimulus, especially the input optical power, could indicate a broken photo-detector. This is a sensible conclusion if the TIA is powered on and drawing a total supply current that closely matches specifications. If this is the case, the transistors of the TIA have not been damaged due to ESD or accidental high-voltages while powering on and setting the DC power supply. If the PD is broken and not transmitting data, only the inherent noise of the TIA will be sensed. Since each part of the optical receiver used its own power supply, it is particularly important that the PD and TIA supply are turned on simultaneously.

This is to ensure that the PD never gets forward biased. Even 1 V of forward bias is enough to generate a large diode current that damages the PD. Since the TIA utilizes a power supply of 2.5 V, it can be assumed that the input of the TIA will never sit at a voltage above 2.5 V. Therefore, the PD's anode, which is directly connected to the power supply, should be set to at least 3 V before turning on the DC power supply. This would guarantee a minimum and maximum PD reverse bias voltage of 0.5 and 3 V, respectively. During testing and troubleshooting, the PD supply voltage should never be set below 2.5 V, despite what the optimal reverse bias might be. Ideally, there would be two extra pads on the TIA die to eliminate the limitation described above: 1) a supply pad that sources the required PD supply and 2) a pad that feeds this supply from the TIA chip to the PD (after using on-chip capacitors on the TIA for filtering).

2) Since the wire-bonding of the dies were outsourced, all connections were done at once. This included the wire-bond that carries the signal from the cathode of the PD to the input of the TIA. Therefore, it was not possible to monitor the input of the TIA beforehand. Resource-permitting, the TIA should be fully wire-bonded (minus its input) to a board or carrier package first. A DC test should be performed to confirm if the die is undamaged. Then, the input pad of the TIA should be measured and recorded with a DC probe. Finally, the PD should be added and wire-bonded to a carrier as a secondary step. In this two-step process, the tester has a more accurate idea of how to bias the PD based on the input of the TIA. The TIA was provided by academic collaborators of the industrial partner INTENGENT Inc., and thus knowledge of the TIA was limited to its test results. Little information was known about the input circuitry, biasing, and whether ESD protection circuits were included on the input or supply pads. Time and resource-permitting, it is important to test the TIA individually before wire-bonding to a PD in this circumstance.

3) Although the wire-bond between the PD and TIA was importantly quantified, the wire-bond supplying power from the board to the PD is also in the RF path of the signal. Unlike the TIA, the PD does not have on-chip capacitors for both supply noise filtering and bypassing the inductive supply path. To circumvent this, discrete capacitors should be placed as close to the PD as possible. One end of the capacitor will have a short wire-bond connected to the PD's anode. Since a discrete capacitor will be much larger than the pads of the PD and TIA, the same end should have a wire-bond going out to the power supply on the PCB or carrier package. The other node of the capacitor needs to be connected to ground with as little inductance as possible. Thus,

an exposed metal ground area should be reserved on the board for die attachment. This way, the bottom of a capacitor will sit directly on ground. As long as the bottom of the PD and TIA dies can be grounded, this solution should pose no problem and help minimize RF signal attenuation by the PD's power supply wire-bonds.

4) In this case, there was no access to the schematics of the TIA. However, it is important to ensure that the on-board bypass capacitors along with the wire-bonds connected to the die's supply pads do not form an LC oscillator. A simple simulation, which includes these off-chip components, with the first stage of a simple TIA or a 50 Ω differential driver should be conducted prior to designing and populating the board.

4.5 Summary

In this chapter, a high-speed photodetector was characterized and interfaced with a TIA following a traditional optical sub-assembly process. This provided quantitative insight into a photodetector's performance and its parasitics. The methodology involved with obtaining the various quantitative parameters is described in Table 4.3. This model will be used to aid the design of the optical receiver front-end with a TIA in the following chapter. Although one design iteration is followed in this thesis based on the PD model characterized in this chapter, electro-photonic design will optimize both PD and TIA parameters in a more iterative fashion that best meets customer requirements at lowest cost.

Table 4.3 Summary of photodetector parameters and the experimental/analytical methodology utilized to extract them.

Measurement parameter	Value	Methodology
Responsivity R	0.73 A/W at 1310 and 1510 nm; 0.4 A/W at 850 nm	Continuous-wave optical measurements. Responsivity = measured current/input optical power (A/W)
Dark current	0.07 nA	Measurement of current from photodetector under no-light conditions (from laser and stray light from visible light sources)
PD pad capacitance C_p	50 fF	Linear extrapolation of total capacitance vs. PD junction area to junction area = 0 cm^2 + confirmation with S-parameter measurements
PD junction capacitance C_j	54 fF	Subtraction of total capacitance from extrapolated pad capacitance + confirmation with S-parameter measurements
Series resistance R_s	8 Ω	Approximated by taking the derivative dV/dI of the photo-detector's current-voltage (IV) curve near the transition point from reverse to forward bias (i.e., the zero-voltage bias point)
Bond-wire inductance L_w	350 pH	Knowledge of PD-TIA chip dimensions + averaging bond-lengths observable from PD-TIA board assembly micrographs. Bond-wire estimated as a semi-circular arc for length calculations and inductance was formulated using the rule of thumb: 1 nH ≈ 1 mm of bond-wire

Chapter 5

5 Front-end system design and analysis

By using the concepts covered in the two previous chapters, this chapter discusses a top-down design approach and methodology for analyzing and quantifying the benefits of monolithic electro-photonic integration versus sub-assembly for next generation optical receiver front-ends.

A TIA connected in negative feedback as in Figure 5.1 on the following page, also known as a shunt-shunt feedback TIA or simply feedback TIA, is used as a fundamental building block for the entirety of the design and analysis process. It is a commonly-used configuration due to its many advantages including, but not limited to the following:

1) The shunt-shunt feedback topology reduces the input and output impedance of the front-end receiver by sensing a voltage at the output and returning a current to the input.

2) The input pole becomes larger in magnitude and, theoretically, higher bandwidths can be achieved. A larger portion of the photo-current from the photodetector is injected into the amplifier as opposed to being dissipated by shunt parasitics. Intuitively, a current divider is formed at the input, where the current source is the photo-current that is divided between its own impedance (including shunt parasitics) and the input impedance of the TIA. The lower the input impedance of the TIA, the higher the current available for current-to-voltage amplification.

3) The output is better suited to drive large loads or serve as the input to a subsequent voltage-to-voltage amplifier. A voltage divider is formed at the output. Large capacitive loading from subsequent stages will present a low impedance AC signal path. Thus, the lower the output impedance of the TIA, the higher the signal available for voltage-to-voltage amplification.

4) With a low input impedance, a time-varying voltage signal applied at the input of the amplifier varies minimally with respect to a time-varying photo-current – keeping the input transistors of the amplifier in their intended range of operation. This is advantageous when high signal swings are present at the input. It also aids with minimizing distortion when linearity is important, such as for modulation schemes like PAM-4.

5) Another clear advantage of this configuration is that the transimpedance element R_f can be increased without worrying about voltage headroom issues.

Figure 5.1 shows a schematic of shunt-shunt feedback TIA, with a transimpedance element R_f, photo-current I_{ph}, a voltage-to-voltage amplifier transfer function $A(s)$, and an input capacitance C_{in}.

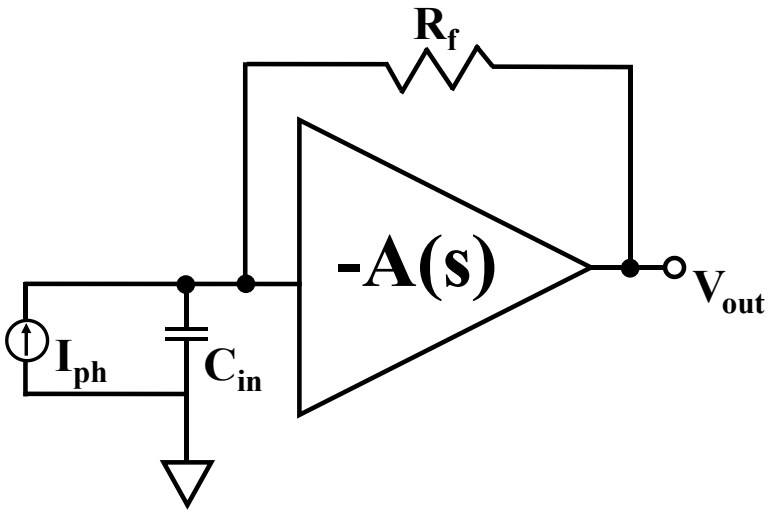

Figure 5.1 A shunt-shunt feedback TIA with an input photo-current I_{ph} and capacitance C_{in}.

5.1 Analytical Modeling

5.1.1 First order

Assuming that the amplifier in the shunt-feedback topology of Figure 5.1 has constant gain A (i.e. an infinite or very large poles), it can be analytically modeled as a first order system with the transfer function described by equation 5.1.

$$\frac{V_{out}}{I_{in}} = \frac{-AR_f}{sC_{in}R_f + A + 1} \tag{5.1}$$

From equation 5.1, it can be seen that the DC to mid-band transimpedance gain is approximately -R_f for A >>1. This is often referred to as the transimpedance element and is an important design variable. It impacts the gain and bandwidth of the receiver front-end as well as the noise performance, as discussed in Chapter 4. The input impedance at DC, also known as the input resistance R_{in}, is easily derived by plugging in s = 0 into equation 5.1 and realizing that V_{out} is simply (V_{in}) x (A). The input resistance and the 3-dB bandwidth is described by equations 5.2 and 5.3, respectively. Recall that this is the bandwidth of the entire closed-loop system and is different from what the open loop poles of the front-end may be.

$$R_{in} = \frac{R_f}{A + 1} \tag{5.2}$$

$$f_{3dB} = \frac{A + 1}{2\pi C_{in}R_f} \tag{5.3}$$

Usually, C_{in} is fixed by the photodetector and photodetector-TIA interface. Therefore, the main design variables that can be fine-tuned by an electronic designer is the transimpedance gain element R_f and amplifier gain A. In the first order approximation of the system, the amplifier gain A required to fulfill a bandwidth requirement scales linearly with the input capacitance, while all other parameters are held constant. Using the design rule of setting the 3-dB bandwidth of the system to 70% of the data rate, amplifier gain A is plotted against input capacitance for various data rates in Figure 5.2. The feedback resistance R_f is set to 300 Ω, which roughly corresponds to a transimpedance gain of 50 dBΩ. In typical receivers, a total transimpedance gain of 70 – 80 dBΩ, including the TIA and post-amplification stages, is required to amplify the bare minimum photo-current that yields a BER of 10^{-12} to a level detectable by clock and data recovery circuits. From a noise and power perspective, it is better to allocate more transimpedance gain at the TIA stage because it reduces the gain requirement of any post-amplification stages. Allocating more gain at the TIA stage proportionally reduces the noise contributed by downstream circuits because the input referred noise, which is compared with the photo-current signal at the input, is reduced

proportionally when gain is increased [25], [58]. With less gain required by the post amplifiers, it becomes easier to design high gain-bandwidth post amplifiers.

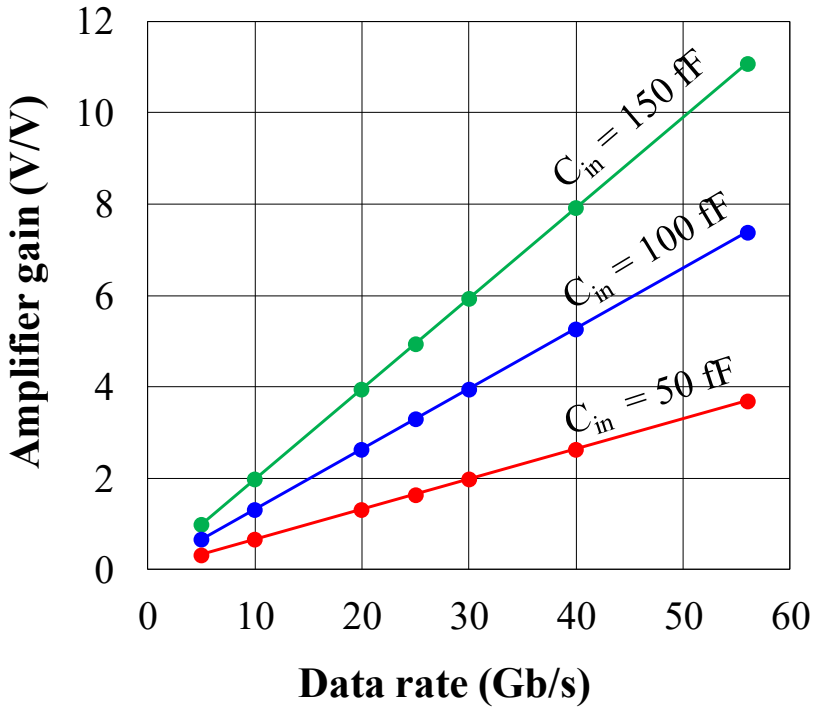

Figure 5.2 Required amplifier gain versus data rate for a first-order shunt-shunt feedback TIA with various input capacitances of 50 fF, 100 fF, and 150 fF.

Although simple, a designer can use the first order approximation of a shunt-feedback topology in the initial design stages of a TIA. For example, a realistic input capacitance of 150 fF necessitates a relatively high TIA gain of at least 10 V/V for 50 Gb/s operation. Thus, power consumption scales as data rates increase. To meet a data rate, the designer must either cascade multiple stages within the feedback loop or select a high-gain technology – SiGe HBTs or InP HBTs. If a CMOS amplifier cannot satisfy this requirement, the designer can still reduce the R_f of 300 Ω to better meet data requirements. Of course, this increases noise and sacrifices sensitivity. This is one of many possible reasons why some literature sources speculate that the sensitivity of CMOS receivers is inferior to that of bipolar technologies as data rates continue to escalate.

Though the input capacitance is fixed by the photodetector and packaging environment, these constraints can be alleviated via monolithic electro-photonic design. When both the photodetector and TIA are co-designed, the designer has the ability to sacrifice detector responsivity via junction capacitance reduction, which saves power via gain reduction, and recover sensitivity by increasing the transimpedance element R_f. In traditional sub-assembly, the TIA design is more constrained.

In terms of meeting the data rate requirement of the application, first-order modeling suggests that single-stage bipolar amplifiers (in III-V or SiGe), whose per-stage gain-bandwidth is typically higher than CMOS, will excel. Despite these constraints, functional 50 Gb/s CMOS TIAs were presented in the literature survey. Of course, some of the reported designs in CMOS did not sub-assemble the TIA with a photodetector and reported measurements and simulations using a single input capacitance of 50 fF. In Figure 5.2, this input capacitance brings down the gain requirement to a mere 3 V/V at 50 Gb/s and, in the proceeding sections, it will be shown that this is feasible for a single-stage, 65 nm CMOS amplifier. The impact of under-estimating the input parasitics will be demonstrated. This will help clarify if CMOS and optical sub-assembly is actually inferior to bipolar technologies and monolithic integration for 50+ Gb/s data rates.

The major limitation of this model is that it assumes the amplifier's internal pole is infinite (i.e. gain A is frequency-independent), which is obviously not the case. Whereas the f_T of InP and SiGe-based transistors have been traditionally higher than CMOS, enabling larger amplifier poles, technology scaling in CMOS can push internal amplifier poles further away. This makes the first order approximation more applicable. Nonetheless, the following section extends this model by assuming a single-pole amplifier.

5.1.2 Second-order

When $A(s)$ is not a constant function, the treatment of the feedback TIA changes:

$$\frac{V_{out}}{I_{in}} = \frac{-A(s)R_f}{sC_{in}R_f + A(s) + 1} \tag{5.4}$$

where the amplifier can be approximated as one-pole system:

$$A(s) = \frac{A_o}{1 + \frac{s}{\omega_o}} \tag{5.5}$$

where A_o is the open-loop, DC gain of the amplifier (V/V) and ω_o is the internal pole (also the open-loop bandwidth) of the amplifier. Now, the transfer function becomes second order, as described by 5.6:

$$\frac{V_{out}}{I_{in}} = -\frac{\frac{A_o \omega_o}{C_{in}}}{s^2 + s\left(\omega_o + \frac{1}{C_{in}R_f}\right) + \frac{(A_o + 1)\omega_o}{C_{in}R_f}} \tag{5.6}$$

The DC transimpedance gain and input resistance, which can be found by plugging s = 0 into expression 5.6, is the same as the first order model. By using second-order dynamics (see 5.7), the damping factor ς and natural frequency ω_n can be extracted from the denominator of 5.6:

$$s^2 + (2\varsigma\omega_n)s + \omega_n \tag{5.7}$$

$$\varsigma = \frac{1}{2}\frac{C_{in}R_f\ \omega_o + 1}{\sqrt{(A_o + 1)\omega_o R_f C_{in}}} \tag{5.8}$$

$$\omega_n = \sqrt{\frac{(A_o + 1\ )\omega_o}{C_{in}R_f}} \tag{5.9}$$

This reveals two interesting revelations about the shunt-feedback topology: its bandwidth can theoretically be higher than what first order modeling suggests and this can even lead to frequency-response gain peaking and overshoot/ringing in the time domain, which resembles inductive behaviour. A damping factor of 1 is desirable in second order systems, but usually a factor of $1/\sqrt{2}$ or 0.7 is suffice [25].

Plotting the damping factor for a single-stage, open-loop amplifier gain A_o of 8 V/V (18 dB) in Figure 5.3, it can be seen that the further the designer places the internal pole of the amplifier, the larger the damping factor. This leads to less the ringing and overshoot in the transient response. The target of 0.7 is shown as a dashed black line.

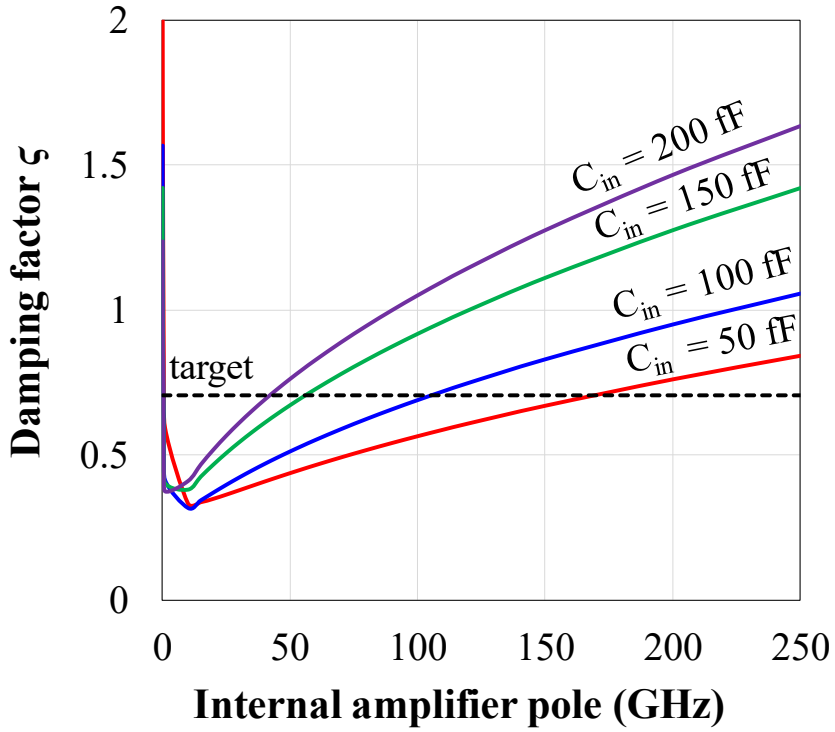

Figure 5.3 Plotting the damping factor of a second-order TIA model over the amplifier's internal pole location for various input capacitances from 50 fF to 200 fF.

By plotting the damping factor vs. the internal pole of the amplifier for various input capacitances, it can be concluded from Figure 5.3 that the relative placement of the internal pole ω_o with respect to the input pole is what dictates whether the 2nd order system will have a sufficient damping factor or not. This is because the input pole is governed by the input capacitance. From equation 5.8, a larger input capacitance (or, smaller input pole) results in a larger damping factor. Therefore, the internal pole ω_o that is required to set equation 5.8 equal to 0.7 is less. This explains why the damping factor crosses the target at a much larger internal pole frequency for an input capacitance of 50 fF compared to input capacitance of 200 fF. Increasing feedback resistance R_f,

which also governs the magnitude of the input pole in a similar fashion to C_{in}, leads to a larger damping factor. Of course, this is at the cost of reduced system bandwidth. The proportionality relationship between ς, R_f, C_{in}, and ω_o is described by expression 5.10:

$$\varsigma \propto \sqrt{R_f C_{in} \omega_o} \tag{5.10}$$

Knowing that the input pole is inversely proportional to R_f and C_{in}, expression 5.11 summarizes the damping factor's dependence on the relative magnitudes of the internal pole ω_o and input pole ω_{in}:

$$\varsigma \propto \sqrt{\frac{\omega_o}{\omega_{in}}} \tag{5.11}$$

Modeling the shunt-shunt feedback topology as a second order system better demonstrates the utility of technology scaling in CMOS, where system stability improves with larger, or faster, poles. High f_T technologies, brought forward by technology scaling in CMOS, lower the total input capacitance and this necessitates a higher internal pole for minimizing overshoot and ringing in the time domain. Fortunately, high f_T technologies also offer these required pole magnitudes.

It can also be seen that designing a photo-detector with less junction capacitance can be beneficial for bandwidth but this comes at the cost of requiring a larger internal amplifier pole, as established by equation 5.11. Even in flip-chip environments, where parasitic inductance is low, the pairing of a low-capacitance, high-speed, off-the-shelf photo-detector with a sub-par might degrade the overall performance of the RX front-end. In the monolithic electro-photonic integration, the co-design of the detector and TIA can minimize undesired behaviour in the time-domain explained by this second-order model.

In the figure below, the damping factor is plotted as a function of amplifier pole location for an input capacitance of 100 fF and two different amplifier gains. A front-end amplifier of a smaller gain (in orange) reduces the magnitude of the internal pole required to meet stability specifications. In this theoretical model, a gain of 4 V/V is sufficient to yield an overall bandwidth

of approximately 30 GHz if the internal amplifier pole is selected to meet the damping factor design target.

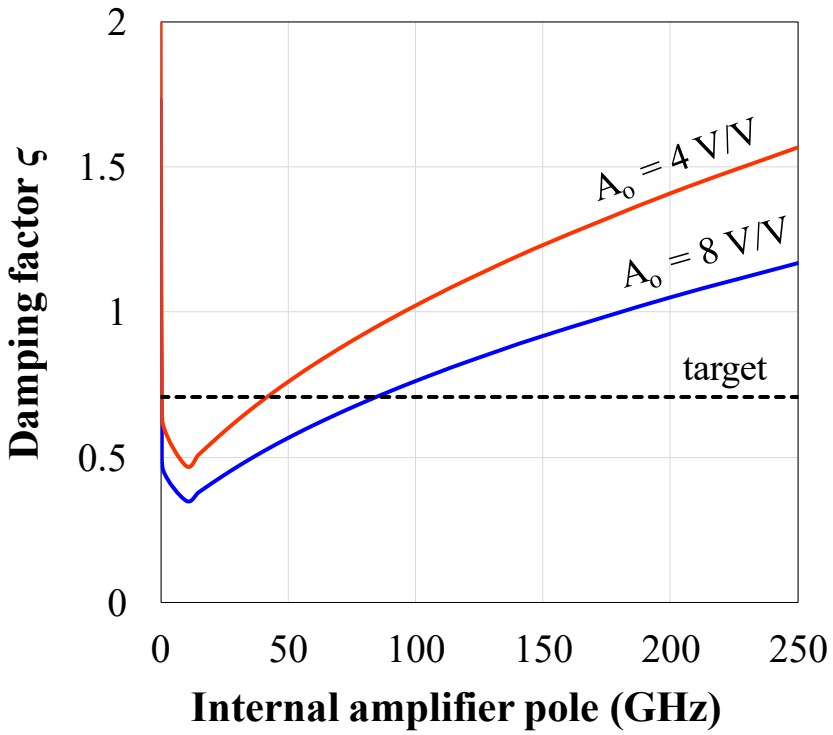

Figure 5.4 Damping factor versus amplifier's internal pole location for two different open-loop amplifier gains and an input capacitance of 100 fF.

5.1.3 Conclusions

In general, shunt-shunt feedback TIA topologies require higher amplifier gain when higher data rates are demanded– owing to a direct relationship between power consumption and data rate. As shown in literature survey, CMOS provides low power consumption primarily due to low supply voltages, but III-V amplifiers may better meet gain – and thus, data rate – requirements.

The larger the pole of the amplifier, the more likely the front-end will maintain stability and bandwidth. This is one of many ways in which technology scaling in CMOS helps and will continue to be contender for next generation data communication products.

In any case, for monolithic integration, reducing the capacitance of the photo-detector is beneficial from a power and technology selection point-of-view. It reduces the gain requirement of the amplifier for a given data rate, which saves power. As demonstrated in Figure 5.4, placement of the internal pole of the amplifier relaxes with lower gain requirements.

5.2 Proposed model and simulations

In this section, a simple model to represent an optical receiver front-end including detector and packaging parasitics is proposed. This model will eventually be validated to a 65 nm CMOS technology node and can be extended to other technology nodes as well. This electrical model requires knowledge of assembly (or packaging) parasitics and an idea of TIA gain per capacitance, which is a characteristic of the technology. It simplifies the evaluation of various technologies against different assembly approaches and vice versa.

In Chapter 4, information related to PD parasitics were obtained through experimental characterization of actual PD dies. In this chapter, these values are used to drive the analysis of the front-end receiver. Bond pad capacitance C_w of the TIA is based on bond pad models in 65 nm CMOS technology and depends on the size of the bond pad of the electronic IC and the metal stack of the technology. The size of the bond pads is limited by the requirements the wire-bond or flip-chip bond. Therefore, this capacitance cannot be easily scaled down. The minimum bond pad size required for wire-bonding or flip-chip bonding is 60 µm x 60 µm, which corresponds to a C_w of approximately 60 fF in 65 nm technology. Note that this includes the capacitance added by electrostatic discharge (ESD) protection circuits.

In order to focus solely on the dynamics of the integration environment between the PD and TIA, the shunt-shunt feedback TIA model shown in Figure 5.5 b) is replaced by its input capacitance C_i and input resistance R_{in} as shown in Figure 5.5 b). This enables the study of the impact of a packaging environment for various values of C_i and R_{in}, which are constituted by the TIA design.

Before designing the TIA and validating this simple model to a known technology, parameters such as packaging inductance L_w and input resistance R_{in} is first swept to identify general trends in signal behaviour at the PD-TIA interface.

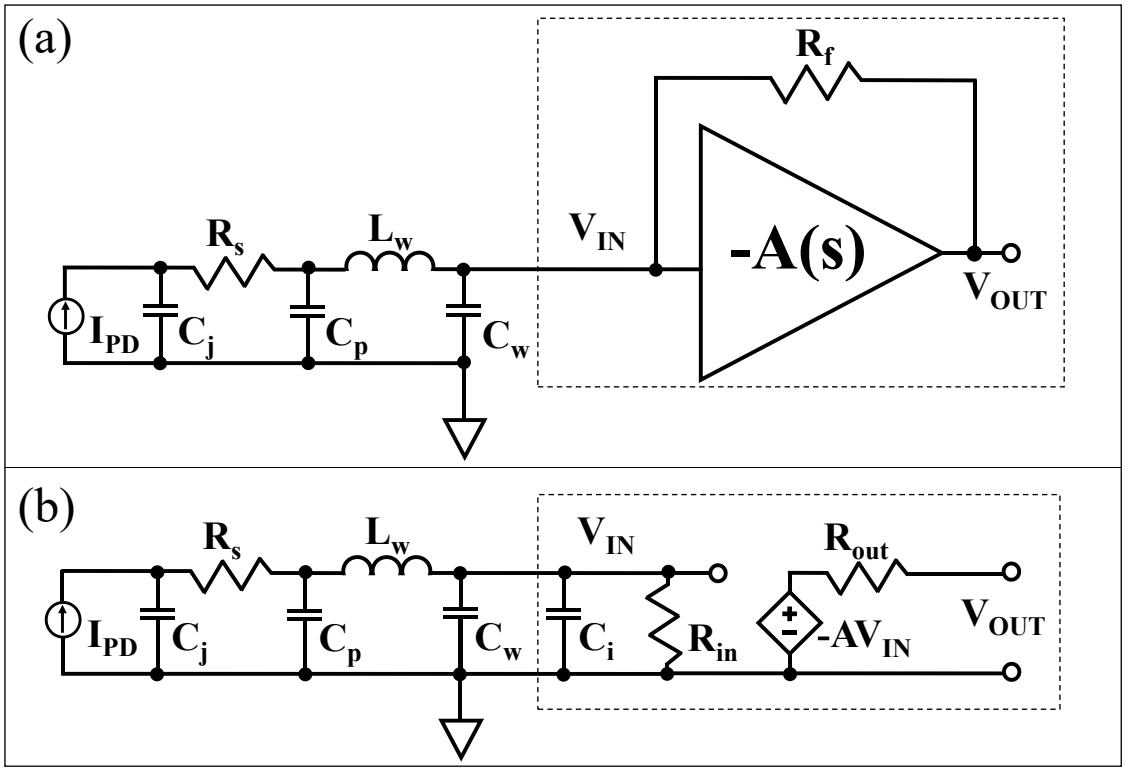

Figure 5.5 Model of a sub-assembled receiver front-end. (a) Shunt-feedback TIA. (b) Configuration simplified by input resistance and input capacitance.

Packaging inductance is swept for a sample input capacitance C_i of 20 fF and input resistance of 50 Ω. In this work, 65 nm CMOS technology is used for the amplifier implementation described in section 5.3. In MOSFETs, the input capacitance is a function of the width of the transistors of the input stage. The larger the width of the transistor, the larger the input capacitance. By doing a sweep of input capacitance versus width for a sample inverter-based amplifier, it was found that about 15 – 35 fF of input capacitance is required for a 25+ Gb/s designs. The transconductance g_m of a MOSFET also increases with width, allowing for greater gain and speed. Thus, there exits an optimal input capacitance or transistor width for a given bandwidth requirement. As will be shown in section 5.3, an input capacitance of about 17 fF results from designing a 50 Gb/s inverter-based TIA built in 65 nm CMOS. Therefore, 20 fF is within range and is selected as a starting point for these preliminary interface simulations.

For a 50 Ω matched input environment, an input resistance of 50 or more ohms is required. Hence, an input resistance of 50 Ω is used to generate Figure 5.6. In the first order TIA model

presented in section 5.3, a transimpedance gain Z_t of 50 dBΩ was allocated to the TIA. Many designs in literature also allocate about 45 to 50 dBΩ of gain to the TIA stage [42], [59], [60]. A gain of 50 dBΩ is 316 Ω. If the transimpedance Z_t is known for a shunt-shunt feedback TIA, the corresponding feedback resistance R_f can be found by knowing the open-loop gain A of the amplifier as described by equation 5.12:

$$R_f = \frac{Z_t(A+1)}{A} \tag{5.12}$$

In many cases, the open-loop gain of the amplifier is unreported in literature. However, a simple table like Table 5.1, where the values of A are swept for given Z_t, can be used to obtain a range of R_f values using equation 5.12 and a range of input resistance R_{in} values using equation 5.2. In Table 5.1, the approximate range of feedback resistance and input resistance values are summarized for an open-loop gain range of 3 - 8 V/V. Coincidentally, an input resistance of 50 Ω is a reasonable value for a TIA that has a typical transimpedance gain of 45 to 50 dBΩ.

Table 5.1 Corresponding feedback and input resistance of a feedback TIA for a range of open-loop gain values

Z_t (dBΩ)	Z_t (Ω)	A	R_f (Ω)	R_{in} (Ω)
45	178	3	237	59
		4	223	45
		5	214	36
		6	208	30
		7	203	25
		8	200	22
50	316	3	421	84
		4	395	79
		5	379	63
		6	369	53
		7	361	45
		8	356	40

Combining measurement results from Chapter 4 (re-summarized in Table 5.2) and reasonable values for the input capacitance and resistance of a typical TIA, frequency-domain and time-domain simulations at the TIA input V_{IN} are plotted in Figure 5.6. The voltage step response at the input of the TIA is simulated by the stepping the photo-detector current from 0 to 100 µA in 1 ps. The assembly environment, which is constituted by electro-photonics, is isolated in this analysis approach. The frequency response characteristics are summarized in Table 5.3.

Table 5.2 Summary of PD-TIA interface parasitics, where all values except TIA bond pad capacitance C_w were measured in Chapter 4

Passive Component	Description	Value
C_j	PD junction capacitance	54 fF
R_s	PD series resistance	8 Ω
C_p	PD bond pad capacitance	50 fF
L_w	Inductance between PD-TIA interface	350 pH
C_w	TIA bond pad capacitance	60 fF

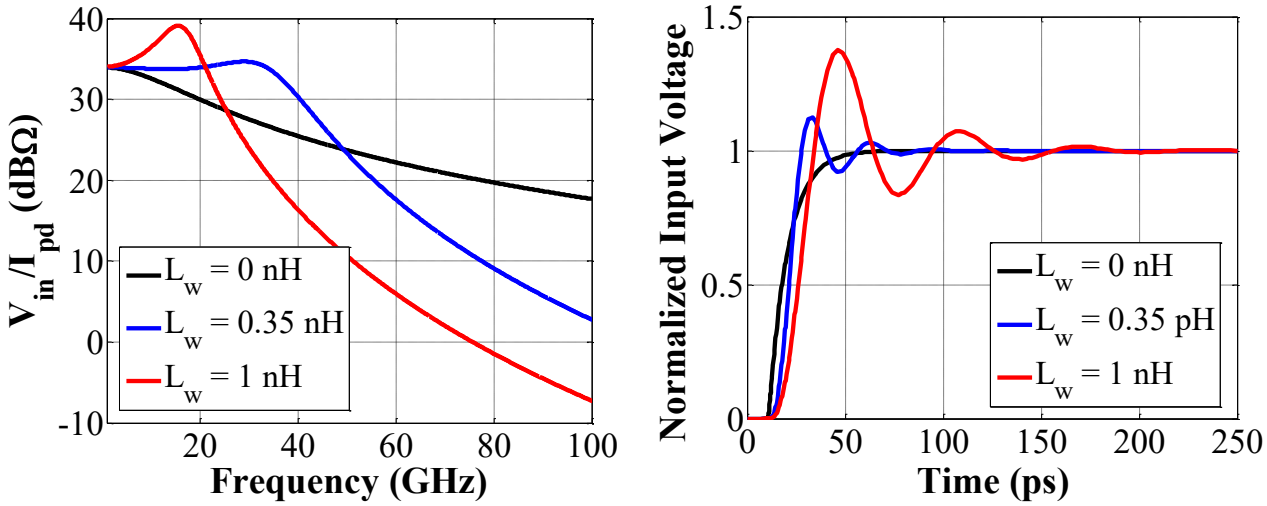

Figure 5.6 Frequency-domain (left) and step response (right) of sweeping L_w for R_{in} = 50 Ω.

Table 5.3 Sweeping package inductance for a TIA input resistance of 50 Ω

L_w (nH)	Bandwidth (GHz)	Peaking magnitude (dB)
0	16	0
0.1	20	3
0.35	39	1
0.5	33	2
1	23	5

In Table 5.3, a wire-bond inductance of 350 pH provides 23 GHz of bandwidth extension with only 1 dB of peaking compared to the zero-inductance case. There is an optimal wire-bond length that extends bandwidth and minimizes peaking. As demonstrated in Figure 5.6, an inductance of 1 nH results in 5 dB of peaking and is 16 GHz lower than the 350 pH case. The optimal inductance depends on the integration parasitics and more specifically, the input resistance of the amplifier. However, in the discussion of packaging in Chapter 4, it was found that the wire-bond inductance cannot be simply selected by the designer. In the PD-TIA packaging experiment, the wire-bond length happened to be 350 pH, which coincidentally is an optimal value for an input resistance and capacitance of 50 Ω and 20 fF, respectively.

Next, the packaging inductance is fixed at the optimal value of 350 pH and the input resistance is swept. If, in fact, the wire-bonding length is predicted to be in this range due to the size of the PD and TIA's pads, the wire-bonder's requirements, etc., then this exercise helps determine the input resistance value that provides an optimally flat frequency response and minimal ringing in the time-domain response. In Figure 5.7, both the frequency and step response at the input of the TIA is plotted. The results of this simulation are summarized in Table 5.4 below.

Table 5.4 Sweeping input resistance of TIA for fixed package inductance of 350 pH

R (Ω)	Bandwidth (GHz)	Peak location (GHz)	Peaking magnitude (dB)
25	40	26	6
50	39	30	1
150	10	30	1

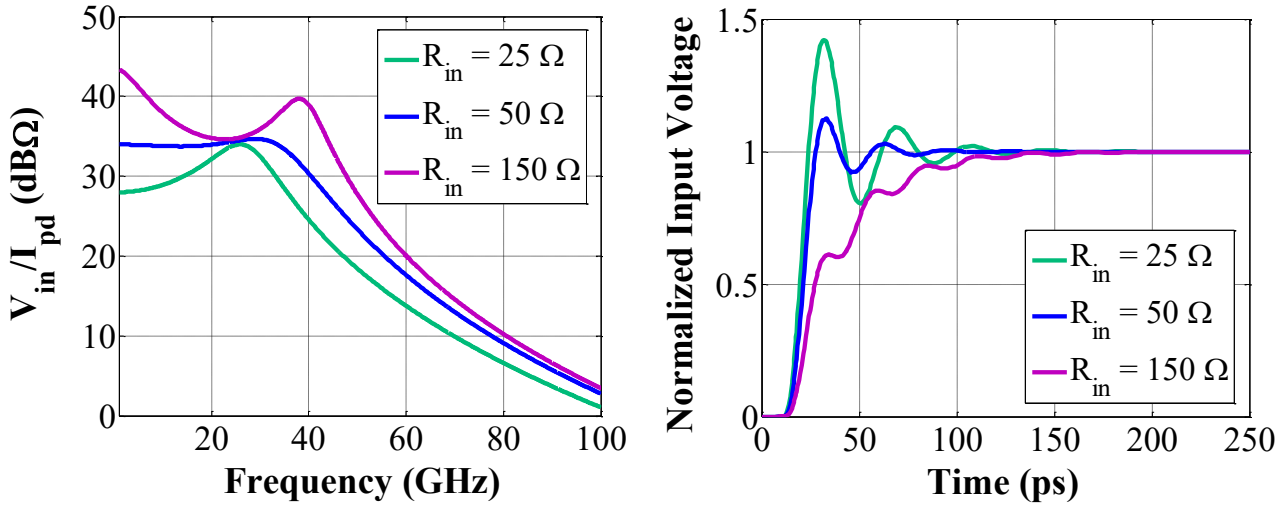

Figure 5.7 Frequency-domain (left) and step response (right) of sweeping R_{in} for L_w = 350 pH.

Figure 5.7 demonstrates the strong dual dependence that stability and bandwidth have on the PD-TIA package inductance L_w and input resistance R_{in} of the TIA. In a simple first order system that consists of only a resistor and capacitor, a low input resistance is always desirable from a speed point-of-view. However, in this more complex environment that represents traditional optical sub-assembly of a detector and TIA, a relatively low input resistance of 25 Ω results in the same bandwidth as the 50 Ω case with an undesirable 6 dB of peaking in the frequency response. Accordingly, the step response exhibits significant overshoot. In this case, the step and frequency responses were taken at the input of the TIA to demonstrate the impact of the optical sub-assembly environment and TIA's input conditions. Depending on the frequency response of the amplifier itself, the ringing may either be nulled or accentuated at the output of the TIA. The input resistance is often set by the noise and data rate requirements of the front-end. However, this analysis suggests that the input resistance of the electronics should be designed based on additional third criteria: minimal gain peaking for a set package inductance. In reality, it can be difficult to accommodate all three aforementioned criteria in selecting an amplifier input resistance, but the electro-photonic designer should be aware of all implications of every design choice.

5.3 Model validation

In this section, the input resistance and capacitance of an amplifier built in 65 nm CMOS is determined. Whether or not this information is enough to at least predict the behaviour at the PD-TIA interface V_{IN} (shown in Figure 5.5 b)) is then determined. In doing so, the validity of this top-down approach to electro-photonic front-end design can be confirmed. These values are then plugged into the simple electrical model shown in Figure 5.5 b). The input capacitance C_i depends on the topology of the TIA, the input transistor size, and the layout. Input resistance is a function of the amplifier's transfer function $A(s)$ and feedback resistance R_f, as described in equation 5.2.

By designing a 50 Gb/s inverter-based amplifier using a 65 nm CMOS electronic design process kit (PDK), the magnitude of the input capacitance Ci and DC input resistance $R_{in,o}$ and input capacitance C_i are identified in Table 5.5. The amplifier gain at DC (A_o), which is determined by dividing V_{OUT} by V_{IN}, along with the feedback resistance R_f is also included in this table. Note that R_f was a design variable that was selected to accommodate this high-bandwidth front-end design. The relationship between the simulated R_{in}, A, and R_f from Table 5.5 follows the theoretical equation 5.2 very closely.

Table 5.5 Important shunt-feedback TIA parameters obtained from 50 Gb/s, 65 nm CMOS inverter-based amplifier.

Parameter	Value
R_f	200 Ω
A_0	2.2 V/V
$R_{in,0}$	63 Ω
C_i	17 fF

To validate the simplified model, the input capacitance and resistance of the sample 65 nm CMOS amplifier is plugged into the simplified model and simulated. In parallel, the PDK-based, transistor-level model implemented using Cadence Design Systems is also simulated. As shown in Figure 5.8 below, the step and frequency response at input V_{IN} are nearly identical in both models. Note that, although the R_{in} is a function of signal frequency, the matching results of Figure 5.8 implies that the internal pole of the amplifier in 65 nm CMOS is high enough, permitting the use of the DC input resistance in the simplified model. Knowing that the amplifier has its own

internal poles and will usually have a capacitive load, it must be noted that the frequency response characteristics at the output of the TIA in this simplified model depicts the best-case scenario. In many cases, however, the PD and its interface with the TIA is the bandwidth-limiting segment of a receiver front-end, making this analysis applicable.

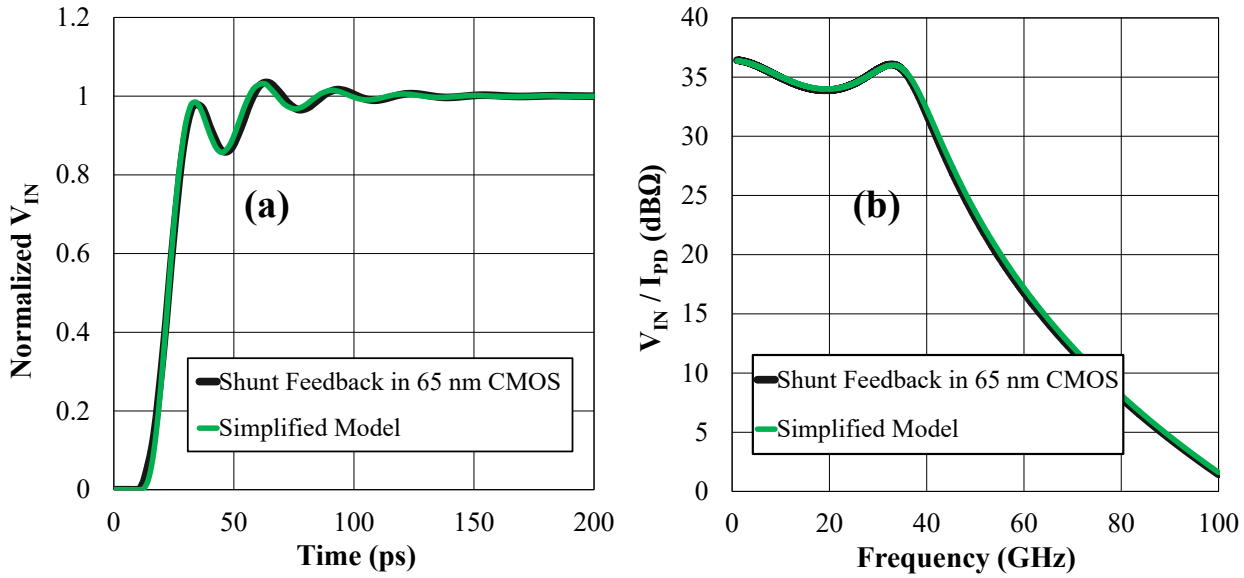

Figure 5.8 Comparing Cadence simulation of a 65 nm receiver front-end with a simplified electrical model, where the CMOS amplifier is replaced by its input resistance and capacitance only. a) Step response. b) Frequency response.

Therefore, modeling the input of a shunt feedback TIA as a single pole system is an adequate predictor of signal behavior at the PD-TIA interface. In fact, it can aid a photonic and electronic designer understand how the physical parameters of their components can optimize the front-end performance. Validated to a specific PDK in this study, the simplified model can be extended to other technologies as well. When using smaller CMOS technology nodes or higher f_T technologies, for example, one can expect a lower input capacitance for the same amplifier gain.

5.4 Impact of integration

The impact of various integration schemes on optical receiver front-ends is investigated in this section. In the technology comparison section of chapter 3, some CMOS TIAs were designed with a single input capacitance of 40 – 50 fF and tested without a photo-detector. Since 50 Gb/s TIA data rates require CMOS technology nodes below 65nm, some of the investigated TIAs were designed for a monolithic integration scenario without the availability of a high-speed detector using the same technology process. It can be argued that major conclusions about the sensitivity of CMOS versus bipolar TIAs cannot be made due to this under-estimation of parasitics reported for CMOS TIAs. In this section, the impact of optical sub-assembly for 50 Gb/s front-ends is assessed in order to validate or invalidate the conclusions of the literature study in chapter 3.

5.4.1 Methodology

A 50 Gb/s CMOS TIA is designed using a single input capacitance of 50 fF, depicting the monolithic scenario encountered in the aforementioned literature survey. By replacing this capacitance with an electrical model more representative of a sub-assembled PD-TIA interface, the impact of packaging is assessed.

5.4.2 Wire-bonding versus flip-chip bonding

Wire-bonding is a common approach for electrically connecting a photo-detector with an optical receiver. Maintaining signal integrity and controlling the length of the wire-bond are some of the major challenges in wire-bonded optical receiver environments. The length of the wire-bond can either aid with extending the bandwidth of the receiver or cause ringing in the time-domain response. Signal integrity issues can include this ringing in the receiver response as well as cross-talk in parallel optical links. Controlling the length of the wire-bond is especially critical in applications that require a certain signal impedance and in the case of differential signal routing.

The inductance of the ball bumps used for flip-chip bonding is much smaller than the inductance from wire-bonds, rendering it favourable for high-frequency applications. A

significantly higher package density can also be achieved with flip-chip bonding as opposed to wire-bonding. Thus, it is a competitive technology for applications such as wireless and optical communications, where integration density can differentiate one product from the other. A major drawback of flip-chip bonding is the cost.

In flip-chip bonding, parasitic capacitance tends to be higher than in wire-bonded environments [61]. This is due to closer spacing between metallic pads as well as between metallic pads and the grounded substrate of one of the chips. In general, the modeling of parasitics in flip-chip and wire-bonded environments should be slightly altered to better predict the performance of the resulting circuits. For example, it has been shown that the quality factor and inductance of on-chip inductors used in RFICs are different in wire-bonded and flip-chip environments [61]. This is an important consideration for TIA topologies that utilize on-chip inductors for bandwidth extension.

Table 5.6 summarizes packaging parasitics values for a flip-chip and wire-bonding scenario [61]–[63]. Flip-chip parasitics are estimated based on simulation and experimental values in literature. As shown in Table 5.6, only a C_j of 50 fF exists for the monolithic scenario used for designing a 50 Gb/s TIA. This table provides all relevant information required for utilizing the simplified electrical model to evaluate the impact of packaging on a sample 50 Gb/s optical receiver front-end.

Table 5.6 Summary of simulation parameters used for three different electro-photonic integration methods of a 50 Gb/s optical receiver front-end.

Parameter	Monolithic	Wire-bond	Flip-chip
C_j	50 fF	54 fF	54 fF
R_s	-	8 Ω	8 Ω
C_p	-	50 fF	50 fF
L_w	-	350 pH	50 pH
C_w	-	60 fF	(60 + 25) fF
R_f	200 Ω	200 Ω	200 Ω
A	2.2 V/V	2.2 V/V	2.2 V/V
R_{in}	63 Ω	63 Ω	63 Ω
C_i	17 fF	17 fF	17 fF

5.4.3 Simulation results

In the Figure 5.9 a), the step response is simulated for all three cases. The monolithic case has the fastest rising time due to minimal parasitics. Sub-assembly via wire-bonding introduces some ringing and overshoot in the step response. In the flip-chip approach, an over-damped response and slow rise time is observed. In the hybrid approach, ringing due to wire-bonding can be seen. Wire-bonding can be beneficial for bandwidth extension, but the wire-bond inductance needs to be optimized for a critically damped response. Whereas the larger inductance introduced by wire-bonding helps extend the bandwidth of the front-end, the lower inductance of a flip-chip bond poorly isolates the PD and TIA capacitance. This is a known phenomenon that results in a larger effective capacitance at the input [64].

To quantify the impact of packaging, the relative impact of parasitics on the frequency response of the system is shown by dividing the output voltage V_{OUT} of the monolithically integrated receiver by the output voltage V_{OUT} of the hybrid integrated receivers in Figure 5.9 b).

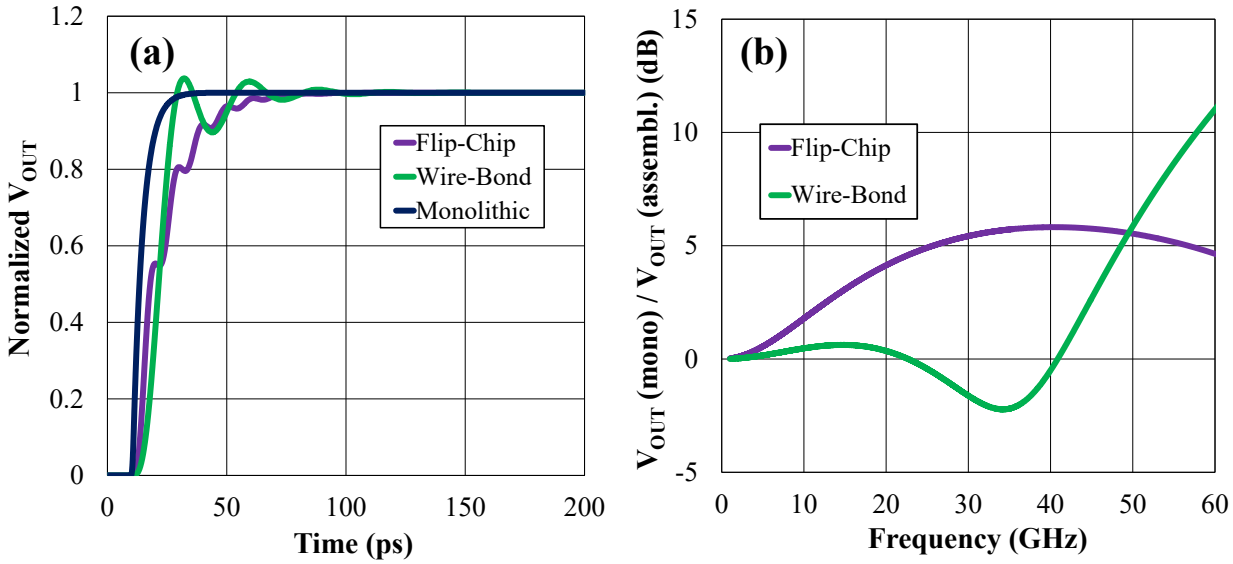

Figure 5.9 Comparing 50 Gb/s receiver front-end with various integration schemes. (a) Step response. (b) Frequency response of a monolithically-integrated PD-TIA divided by the same PD-TIA sub-assembled via wire-bonding or flip-chip means.

In Figure 5.9 b), the ratio of the frequency responses of the monolithic and wire-bonded front-ends is comparable up to 25 GHz. Peaking is observed at around 35 GHz occurs, which provides bandwidth extension and a boost in the output voltage swing for the wire-bonded front-end. At even higher frequencies (above 40 GHz), the wire-bond inductor attenuates the signal and the monolithic approach becomes more suitable. Whereas the larger inductance introduced by wire-bonding helps extend the bandwidth of the front-end, the lower inductance of a flip-chip bond poorly isolates PD and TIA capacitance. This phenomenon, along with a 25 fF increase in parasitic capacitance, results in a larger effective capacitance at the PD-TIA interface, contributing to the slow rise time in Figure 5.9 a) and the poor frequency response shown in Figure 5.9 b). The voltage swing at the output of a flip-chip integrated front-end becomes half that of a monolithically-integrated front-end when the signal frequency is above 30 GHz. Thus, a front-end using a flip-chip scheme must sacrifice input referred noise by using a smaller feedback resistor, move to a higher f_T technology to obtain a higher gain per input capacitance, or both, in order to maintain 50 Gb/s data rate. However, C_j, C_p, and C_w cannot be scaled as easily. The flip-chip method is thus the least sensitive of the three integration methods explored in this thesis, albeit more compact than wire-bonding.

To further study the impact of optical sub-assembly on a high-speed TIA, a 50 Gb/s PRBS-10^{-7} NRZ data stream was generated and transmitted through the same three front-ends described in Figure 5.9. The resulting eye diagrams are illustrated in Figure 5.10. A ‘1’ level of 50 µA and a ‘0’ level of 0 µA was chosen for input photo-current from the PD. Since the input resistance of the TIA is 63 Ω and open-loop voltage gain A of the amplifier is - 2.2 V/V, the peak-to-peak eye height, as shown in Figure 5.10, at the output of the TIA is 50 µA x 63 Ω x 2.2 V/V = 7 mV. To no surprise, the severely reduced bandwidth of the flip-chipped receiver front-end results in poor-quality, unresolvable, eye (far-right in purple). The rise and fall times are visibly slow. Although the wire-bond case (middle in green) has more jitter than the monolithic front-end (far-left in blue), the eye opening (labelled) in mV is comparable and the eye appears to be very symmetric. In both cases, there is a clear region in the middle, where the eye opening is labelled, where the eye is most open and this is also the optimal point at which the data should be sampled for minimum BER.

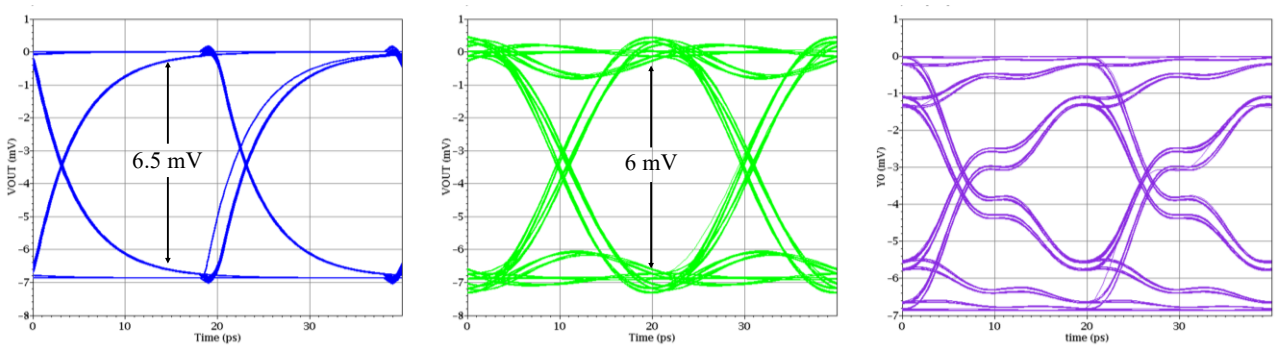

Figure 5.10 50 Gb/s eye diagrams for a 50 Gb/s TIA integrated with a PD in three ways: monolithic (far-left), wire-bond (middle), and flip-chip (far-right).

This analysis shows that designing a 50 Gb/s TIA with minimal parasitics (i.e. monolithic integration) and later wire-bonding it to a PD, does not sacrifice the bandwidth of the front-end, regardless of additional parasitics. Thus, the results of the technology comparison of TIAs in CMOS, SiGe, and InP does not change. All three semiconductor technology platforms are in fact capable of achieving 50 Gb/s data rates - and possibly higher due to technology scaling in CMOS - at comparable sensitivities. Yet, it is critical to note that if these same TIAs were to be flip-chip bonded to a high-speed PD, the bandwidth would suffer – altering the results of the literature survey.

Using on-chip inductors at the input of the TIA or output of the photodetector, in lieu of or in addition to wire-bond, is also a commonly utilized bandwidth extension technique [31], [40]. This approach, along with using low-capacitance waveguide photodetectors discussed in Chapter 2, may enable flip-chip techniques to keep up with the demand for continually increasing data transfer systems. This would allow designers to more readily control and utilize the benefits of inductive peaking demonstrated in Figures 5.9 and 5.10, while removing the added footprint of wire-bond assembly. Although flip-chip techniques would benefit most from waveguide photodetectors whose capacitance can be on the order of 5 times less than the PDs considered in this work, an on-chip inductor on the PD or TIA will likely be required to compensate for the parasitic pad capacitances outlined in Table 5.6 [29], [31]. Of course, the ideal combination for maximizing speed would be to monolithically integrate a SiGe waveguide photodetector with a SiGe BiCMOS process where designers have access to high-speed SiGe heterojunction bipolar transistors.

In addition to added footprint, wire-bonded front-ends are still limited by two factors: 1) significant signal attenuation at frequencies above 40 GHz for realistic wire-bond lengths, and 2) complexity of optimizing for a critically-damped response. The cross-over frequency at which monolithic front-ends are better would be lower if a larger wire-bond inductance is used. In the other hypothetical limit of reducing the wire-bond inductance as much as possible, the frequency response approaches that of the flip-chip case – undesirable at 50 Gb/s and above. At realistic inductance values of 200 – 400 pH, the frequencies at which signal attenuation compromises the sensitivity of optical receivers, is currently being approached. This drives the need for monolithically-integrated solutions.

Chapter 6

6 Conclusions

6.1 Conclusion

Electro-photonic integration, which involves both the co-design and possibly, monolithic integration of electronic and photonic functions was introduced and investigated. Although monolithic integration of optical transceivers was more frequently explored in the 1980s and 1990s with little commercial success, the current state-of-the-art was explored to re-evaluate its benefits as single-lane data rate demands rise.

Wire-bonding a photodetector to a TIA results in an area penalty of at least 0.25 mm^2. Considering a single transceiver module may contain multiple monolithic receivers and data centers employ many modules, the combined area savings from reduced form factor can minimize cooling requirements.

Along with reduced form factor, it was demonstrated that the performance of traditionally-assembled receiver front-ends are becoming limiting as single lane data rates approach 50 Gb/s and above. In Chapter 2, successfully demonstrated monolithic solutions were presented in silicon, SiGe, and III-V substrates. SiGe and III-V processes can be utilized to fabricate fully monolithic optical receivers for the three main communication bandwidths of 850 nm, 1310 nm, and 1550 nm. However, a broadband anti-reflective coating must be designed for these receivers to be readily usable for all wavelength applications. It was shown, for example, that commercial detectors with an anti-reflective coating optimized for long wavelengths can lose up to 3 dB in

sensitivity when used for short wavelengths. This loss comes from the coating alone as the detectors have comparable theoretical sensitivities at all wavelengths. To entirely utilize the CMOS process without costly modifications to the fabrication process, all-silicon detectors must be designed. Due to silicon's absorption properties, all-silicon detectors will be limited to short wavelength applications.

The current state-of-the-art suggests that front-end optical receivers (excluding the PD) built in SiGe, CMOS, and III-V technology platforms have comparable sensitivity and power consumption. However, many literature sources do not report the experimental results of the receiver's performance after assembling and testing it with a photodetector that receives optical data. As such, an electrical model that captured the PD, packaging environment, and a 50 Gb/s electronic receiver in a known CMOS technology node was demonstrated and simulated in pursuance of validating or invalidating the results of the literature survey. It was confirmed that there is no major performance penalty in excluding the photodetector and packaging environment, if the PD were to be wire-bonded to the TIA for eventual testing. With the wire-bond isolating the PD's capacitance and even providing bandwidth extension, the performance of a TIA with a single input capacitance (resembling the monolithic scenario depicted by many literature sources) and a fully-assembled PD + TIA is comparable up to 40 GHz for an inductance of 350 pH. A higher inductance would reduce this break-even point to lower frequencies, while a lower inductance would more poorly isolate the PD's capacitance. The signal attenuation caused by the wire-bond at 50 GHz is so high that, if the same TIA were to be monolithically integrated to a similar PD, it would have four times the signal swing at its output. Thus, while it was found that most technology platforms offer comparable performance advantages in this current age, the real limitation lies in the assembly of photonic and electronic devices.

6.2 Future Work

The input capacitance C_{in} and gain A of a voltage-to-voltage amplifier is proportional to the f_T of the technology used to design it. High f_T technologies are expected to offer more A per C_{in}. Since the front-end of the receiver often determines the bandwidth and sensitivity of the entire optical receiver, the electrical model should be extended to other technologies as well. Ideally, typical and maximum values of C_{in} and A for different technologies should be stored in a look-up table, which

designers can then use to formulate quick estimates of the technology node that best fits the application at hand. For example, if the application requires very low noise, the designer will choose a technology platform that allows for maximum optimization of the transimpedance element R_f. Since the input resistance of shunt-feedback TIAs is proportional to $R_f / (1 + A)$, this application would perform best with the higher f_T technologies in the table. If a similar approach is taken for photodetector parasitics or transit-times (if speed is not limited by parasitics), the receiver designer can save the time and heavy costs associated with accessing and relying on foundry information prior to a design. This is how electro-photonic co-design could reduce some of the overhead and inaccuracies in the optical receiver design process. For optical transceiver module vendors, it could be economically compelling to purchase the PD-TIA and even digital components in a complete die or package created by one design house. It could also help build trust and credibility in the electro-photonic design house because PD and TIA specifications that traditionally came from two different sources is now provided by one source as a combined set of system-level specification that holds more practical value.

In order to put the claims of monolithic integration into practise, a prototypical receiver front-end with a PD and TIA should be designed using a foundry with a track record of fabricating both photonic and electronic devices. For example, the InP detectors investigated in this study were fabricated by Global Communication Semiconductors, who also offer an InP HBT process. The overhead associated with trying to accommodate these similar processes into one as well as the quantitative advantages of this prototype should be compared with the theoretical and simulation-based investigation presented here. By using the latest available technology for data rates above 50 Gb/s, this exercise would re-visit monolithic electro-photonic integration in today's context.

References

[1] I. A. Young *et al.*, “Optical I/O Technology for Tera-Scale Computing,” *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 235–248, 2010.

[2] A. Rylyakov, “High Speed Circuits for Short Reach Optical Communications,” in *Optical Fiber Communication (OFC) Conference*, 2015, pp. 1–34.

[3] B. Razavi, “Introduction to Optical Communications,” in *Design of Integrated Circuits for Optical Communications*, 2nd ed., Hoboken, NJ: Wiley, 2012, pp. 3–4.

[4] S. P. Voinigescu, T. O. Dickson, R. Beerkens, I. Khalid, and P. Westgaard, “A Comparison of Si CMOS, SiGe BiCMOS, and InP HBT Technologies for High-Speed and Millimeter-Wave ICs,” in *Digest of Papers. 2004 Topical Meeting onSilicon Monolithic Integrated Circuits in RF Systems, 2004.*, 2004, pp. 111–114.

[5] P. J. Zampardi, “Will CMOS Amplifiers Ever Kick-GaAs?,” in *Proceedings of the Custom Integrated Circuits Conference*, 2010, pp. 1–4.

[6] L. M. Lunardi and S. Chandrasekhar, “Integrated p-i-n / HBT Photoreceivers for Optical Communications,” in *International Electron Devices Meeting*, 1996, pp. 3–6.

[7] A. C. Carusone, H. Yasotharan, and T. Kao, “CMOS Technology Scaling Considerations for Multi-Gbps Optical Receivers with Integrated Photodetectors,” *IEEE J. Solid-State Circuits*, vol. 46, no. 8, pp. 1832–1842, 2011.

[8] J. J. G. M. Van Der Tol, Y. S. Oei, U. Khalique, R. Ntzel, and M. K. Smit, “InP-based photonic circuits: Comparison of monolithic integration techniques,” *Prog. Quantum Electron.*, vol. 34, no. 4, pp. 135–172, 2010.

[9] D. Neamen, “Extension to Three Dimensions,” in *Semiconductor Physics and Devices*, 4th ed., New York, NY: McGraw-Hill Education, 2011, pp. 83–85.

[10] S. Kasap, “LED Materials and Structures,” in *Optoelectronics and Photonics: Principles and Practices*, 2nd ed., Upper Saddle River, NJ: Pearson, 2012, p. 237.

[11] S. Kasap, “Extrinsic Semiconductors,” in *Optoelectronics and Photonics: Principles and Practices*, 2nd ed., Upper Saddle River, NJ: Pearson, 2012, p. 190.

[12] N. Y. Li *et al.*, “High-Performance 850 nm VCSEL and Photodetector Arrays for 25 Gb/s

Parallel Optical Interconnects," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, 2010, pp. 2–4.

[13] F. Kish *et al.*, "From Visible Light-Emitting Diodes to Large-Scale III-V Photonic Integrated Circuits," *Proc. IEEE*, vol. 101, no. 10, pp. 2255–2270, 2013.

[14] T. P. Pearsall, "$Ga_{0.47}In_{0.53}As$: A Ternary Semiconductor for Photodetector Applications," *IEEE J. Quantum Electron.*, vol. 16, no. 7, pp. 709–720, 1980.

[15] T. P. Pearsall and J. P. Hirtz, "The Carrier Mobilities in $Ga_{0.47}In_{0.53}As$ Grown by Organo-Metallic CVD and Liquid-Phase Epitaxy," *J. Cryst. Growth*, vol. 54, no. 1, pp. 127–131, 1981.

[16] K. Ng and S. Sze, *Physics of Semiconductor Devices*, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2006.

[17] E. Sackinger, *Broadband Circuits for Optical Fiber Communication*. Hoboken, NJ: John Wiley & Sons, 2007.

[18] R. G. Smith and S. D. Personick, *Semiconductor Devices for Optical Communication*, 2nd ed. New York, NY: Springer-Verlag, 1982.

[19] C. Hu, *Modern Semiconductor Devices for Integrated Circuits*, 1st ed. Pearson, 2009.

[20] J. Silc, B. Robic, and T. Ungerer, *Process Architecture: From Dataflow to Superscalar and Beyond*. Heidelberg: Springer, 1999.

[21] W. Sansen, *Analog Design Essentials*. Dordrecht: Springer, 2006.

[22] N. Weimann *et al.*, "InP DHBT Circuits: From Device Physics to 40 Gb/s and 100 Gb/s Transmission System Experiments," *Bell Labs Tech. J.*, vol. 14, no. 3, pp. 43–62, 2009.

[23] J. McMacken, S. Nedeljkovic, J. Gering, and D. Halchin, "HBT Modeling," *IEEE Microwave Magazine*, pp. 2–28, Apr-2008.

[24] P. Gray, P. Hurst, S. Lewis, and R. Meyer, *Analysis and Design of Analog Integrated Circuits*. Hoboken, NJ: John Wiley & Sons, 2009.

[25] B. Razavi, "Transimpedance Amplifiers," in *Design of Integrated Circuits for Optical Communications*, 2nd ed., Hoboken, NJ: Wiley, 2012, pp. 62–129.

[26] M. Mokhtari *et al.*, "InP-HBT Chip-Set for 40-Gb/s Fiber Optical Communication Systems Operational at 3 V," *IEEE J. Solid-State Circuits*, vol. 32, no. 9, pp. 1371–1383, 1997.

[27] M. Bitter, R. Bauknecht, W. Hunziker, and H. Melchior, "Monolithic InGaAs-InP p-i-n/HBT 40-Gb/s Optical Receiver Module," *IEEE Photonics Technol. Lett.*, vol. 12, no. 1, pp. 74–76, 2000.

[28] G. Agrawal, *Lightwave Technology: Components and Devices*, 4th ed. Wiley, 2004.

[29] J. Michel, J. Liu, and L. C. Kimerling, "High-performance Ge-on-Si Photodetectors," *Nat. Photonics*, vol. 4, no. 8, pp. 527–534, 2010.

[30] R. Swoboda and H. Zimmermann, “11Gb/s Monolithically Integrated Silicon Optical Receiver for 850nm Wavelength,” in *International Solid-State Circuits Conference*, 2006.

[31] D. Kucharski, D. Guckenberger, G. Masini, S. Abdalla, J. Witzens, and S. Sahni, “10Gb/s 15mW Optical Receiver with Integrated Germanium Photodetector and Hybrid Inductor Peaking in 0.13μm SOI CMOS Technology,” *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.*, vol. 53, no. 10, pp. 360–361, 2010.

[32] G. Masini, L. Colace, G. Assanto, H. C. Luan, and L. C. Kimerling, “High-Performance p-i-n Ge on Si Photodetectors for the Near Infrared: From Model to Demonstration,” *IEEE Trans. Electron Devices*, vol. 48, no. 6, pp. 1092–1096, 2001.

[33] F. Aznar, S. Celma, and B. Calvo, “Introduction,” in *CMOS Receiver Front-ends for Gigabit Short-Range Optical Communications*, New York, NY: Springer, 2013, pp. 12–13.

[34] L. Zimmermann *et al.*, “Monolithic Integration of Photonic Devices in SiGe BiCMOS,” in *11th International Conference in Group IV Photonics*, 2014, pp. 102–103.

[35] C. H. Jan *et al.*, “RF CMOS Technology Scaling in High-k/Metal Gate Era for RF SoC (System-on-Chip) Applications,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2010, pp. 604–607.

[36] M. Khater *et al.*, “SiGe HBT Technology with fmax/fT = 350/300 GHz and Gate Delay Below 3.3 ps,” in *IEDM Technical Digest. IEEE International Electron Devices Meeting*, 2004, pp. 247–250.

[37] M. Urteaga, R. Pierson, P. Rowell, V. Jain, E. Lobisser, and M. J. W. Rodwell, “130nm InP DHBTs with ft >0.52THz and fmax >1.1THz,” in *Device Research Conference - Conference Digest, DRC*, 2011, pp. 281–282.

[38] M. Schwartz, “Performance of Communication Systems: Limitations Due to Noise,” in *Information Transmission, Modulation, and Noise*, 4th ed., New York, NY: McGraw-Hill, 1990, pp. 427–428.

[39] M. Donhowe, R. Martin, and F. Peters, “Specifying Optical Modulation Amplitude Instead of Extinction Ratio,” York, 1999.

[40] J. Jin and S. S. H. Hsu, “A 40-Gb/s Transimpedance Amplifier in 0.18-μm CMOS Technology,” *IEEE J. Solid-State Circuits*, vol. 43, no. 6, pp. 1449–1457, 2008.

[41] C. Kromer, G. Sialm, D. Erni, H. Jäckel, T. Morf, and M. Kossel, “A 40 Gb/s Optical Receiver in 80-nm CMOS for Short-Distance High-Density Interconnects,” *2006 IEEE Asian Solid-State Circuits Conf. ASSCC 2006*, pp. 395–398, 2006.

[42] S. Bashiri, C. Plett, J. Aguirre, and P. Schvan, “A 40 Gb/s Transimpedance Amplifier in 65 nm CMOS,” *Proc. 2010 IEEE Int. Symp. Circuits Syst.*, pp. 757–760, 2010.

[43] B. Moeneclaey *et al.*, “A 40-Gb/s Transimpedance Amplifier for Optical Links,” *IEEE Photonics Technol. Lett.*, vol. 27, no. 13, pp. 1375–1378, 2015.

[44] A. Awny *et al.*, “A 40 Gb/s Monolithically Integrated Linear Photonic Receiver in a 0.25

μm BiCMOS SiGe:C Technology," *IEEE Microw. Wirel. Components Lett.*, vol. 25, no. 7, pp. 469–471, 2015.

[45] K. Kobayashi, "An InP HBT Common-Base Amplifier with Tunable Transimpedance for 40 Gb/s Applications," in *24th Annual Technical Digest Gallium Arsenide Integrated Circuit (GaAs IC) Symposium*, 2002, pp. 155–158.

[46] K. Kobayashi, "State-of-the-Art 60 GHz, 3.6 K-Ohm Transimpedance Amplifier for 40 Gb/s and Beyond," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, 2003, pp. 55–58.

[47] H. Fukuyama *et al.*, "Photoreceiver Module Using an InP HEMT Transimpedance Amplifier for over 40 Gb/s," *IEEE J. Solid-State Circuits*, vol. 39, no. 10, pp. 1690–1696, 2004.

[48] R. Ding, Z. Xuan, T. Baehr-Jones, and M. Hochberg, "A 40-GHz Bandwidth Transimpedance Amplifier with Adjustable Gain-Peaking in 65-nm CMOS," in *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2014, pp. 965–968.

[49] S. G. Kim *et al.*, "A 50-Gb/s Differential Transimpedance Amplifier in 65nm CMOS Technology," in *2014 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, 2014, pp. 357–360.

[50] S. Kudszus, A. Shahani, S. Pavan, D. K. Shaeffer, and M. Tarsia, "A 46-GHz Distributed Transimpedance Amplifier Using SiGe Bipolar Technology," *IEEE MTTS Int. Microw. Symp. Dig. 2003*, vol. 2, pp. 1387–1390, 2003.

[51] O. Using and I. Ingaas, "40-GHz Transimpedance Amplifier With Differential Bipolar Transistors," *Techniques*, vol. 38, no. 9, pp. 1518–1523, 2003.

[52] M. H. Nazari and A. Emami-Neyestanak, "A 24-Gb/s Double-Sampling Receiver for Ultra-Low-Power Optical Communication," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 344–357, 2013.

[53] M. Oehme, J. Werner, E. Kasper, M. Jutzi, and M. Berroth, "High Bandwidth Ge p-i-n Photodetector Integrated on Si," *Appl. Phys. Lett.*, vol. 89, no. 7, pp. 1–4, 2006.

[54] G. B. Kromann, R. D. Gerke, and W. W. X. Huang, "A Hi-Density C4/CBGA Interconnect Technology for a CMOS Microprocessor," *IEEE Trans. Components Packag. Manuf. Technol. Part B*, vol. 19, no. 1, pp. 166–173, 1996.

[55] M. Beals *et al.*, "Process Flow Innovations for Photonic Device Integration in CMOS," *Proc. SPIE*, vol. 6898, no. February, pp. 689804-689804–14, 2008.

[56] C. Gunn, "CMOS Photonics for High-Speed Interconnects," *IEEE Micro*, vol. 26, no. 2, pp. 58–66, 2006.

[57] S. Assefa *et al.*, "A 90nm CMOS Intergated Nanophotonics Technology for 25Gbps WDM Optical Communications Applications," in *2012 International Electron Devices Meeting*, 2012, p. 33.8.1-33.8.3.

[58] B. Razavi, *RF Microelectronics*, 2nd ed. Castleton, NY: Prentice Hall, 2011.

[59] J. S. Weiner *et al.*, "SiGe Differential Transimpedance Amplifier with 50-GHz Bandwidth," *IEEE J. Solid-State Circuits*, vol. 38, no. 9, pp. 1512–1517, 2003.

[60] J. Verbrugghe *et al.*, "Multichannel 25 Gb/s Low-Power Driver and Transimpedance Amplifier Integrated Circuits for 100 Gb/s Optical Links," *J. Light. Technol.*, vol. 32, no. 16, pp. 2877–2885, 2014.

[61] G. A. Lee, M. Megahed, F. M. Rotella, and F. De Flaviis, "Analysis of RF flip-chip on-chip inductance," in *2002 32nd European Microwave Conference, EuMC 2002*, 2002.

[62] D. Staiculescu, A. Sutono, and J. Laskar, "Wideband Scalable Electrical Model for Microwave/Millimeter Wave Flip Chip Interconnects," *IEEE Trans. Adv. Packag.*, vol. 24, no. 3, pp. 255–259, 2001.

[63] S. Saeedi, S. Menezo, G. Pares, and A. Emami, "A 25 Gb/s 3D-Integrated CMOS / Silicon-Photonic Receiver for Low-Power High-Sensitivity Optical Communication," *J. Light. Technol.*, vol. 34, no. 12, pp. 2924–2933, 2016.

[64] C. Sige, B. Transimpedance, C. Li, S. Member, and S. Palermo, "A Low-Power 26-GHz Transformer-Based Regulated Cascode SiGe BiCMOS Transimpedance Amplifier," *IEEE J. Solid-State Circuits*, vol. 48, no. 5, pp. 1264–1275, 2013.