INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

COMPONENTS FOR A LOW-COST INTEGRATED SILICON OPTICAL RECEIVER

**BY
RYAN MACDONALD, B.SC./M.ENG.**

**A thesis submitted to
the Faculty of Graduate Studies
in partial fulfillment of
the requirements for the degree of**

Doctor of Philosophy

**Carleton University
Ottawa, Ontario
September 7, 2001
© copyright
2001, Ryan P. MacDonald**

**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-71940-5

Canada

The undersigned recommend to
the Faculty of Graduate Studies
acceptance of the thesis

**COMPONENTS FOR A LOW-COST INTEGRATED SILICON
OPTICAL RECEIVER**

Submitted by
Ryan P. MacDonald, B.Sc./M.Eng.

in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Dr. M. S. Nakhla
Chair,
Department of Electronics

Dr. B. A. Syrett
Thesis Co-supervisor

Dr. N. G. Tarr
Thesis Co-supervisor

Dr. Siegfried Janz
External Examiner

Carleton University
January, 2002

ABSTRACT

This thesis describes the design and fabrication of two components of an optical receiver: a novel silicon photodetector and a novel silicon amplifier circuit. Interdigitated metal-semiconductor-metal (MSM) photodetectors were fabricated by depositing metal contacts on top of a 2 μm thick layer of polycrystalline silicon (polysilicon). These detectors have a -3 dB bandwidth of 750 MHz and a responsivity of 0.13 A/W at 860 nm. The bandwidth is more than twice that reported for conventional silicon MSM photodetectors as a result of the thin absorbing layer made possible by the higher optical absorption of polysilicon. A simple fabrication process that is compatible with standard VLSI processes, together with good performance characteristics, make this an ideal detector for integrated optoelectronic receivers for use in short distance, parallel optical data links. As well, a 1x16 amplifier array fabricated in silicon bipolar technology is described with application for parallel optical data links. The amplifiers are limiting amplifiers of the transimpedance-transadmittance type, and were designed for low power consumption. Measured results for operation at 622 Mbit/s are given.

ACKNOWLEDGEMENTS

Ryan P. MacDonald would to thank his supervisors Dr. B. A. Syrett and Dr. N. G. Tarr of Carleton University for their guidance and support throughout this project. At the National Research Council the help of Dr. J. Chrostowski and Dr. S. A. Boothroyd was invaluable. The help of Dr. P. Myslinski, Dr. J. Hong, and Ross Misner, all of the Photonics Systems lab of the National Research Council, must also be recognized. Thanks is also due to Lyall Berndt and Carol Adams of Carleton's integrated circuit fabrication facility for helping to develop the process for fabricating the polysilicon MSM photodiode.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
Chapter 1 INTRODUCTION	1
1.1 Thesis Motivation	1
1.2 Why Use Optics	7
1.3 Types of Optical Interconnections	12
1.4 Thesis Contributions	13
1.5 Thesis Outline	17
Chapter 2 OPTICAL DETECTION	19
2.1 Photons in Semiconductors	19
2.2 Properties of Semiconductor Photoconductors	31
Chapter 3 METAL-SEMICONDUCTOR-METAL PHOTODIODES.....	44
3.1 Theory of the MSM diode	44
3.2 Photoconducting Materials	59
Chapter 4 DETECTOR STUDY AND EVALUATION	78
4.1 Introduction	78
4.2 Device Fabrication	86
4.3 Device Model and Simulation	91
4.4 Experimental Results	98
4.5 Additional Simulations	106
Chapter 5 OPTICAL RECEIVER	112
5.1 Introduction	112
5.2 Receiver Fabrication Technology	113
5.3 Receiver Sensitivity	114
5.4 Wideband Amplifiers	129
5.5 Integrated Receiver Design	139
Chapter 6 INTEGRATION OF THE POLYSILICON MSM PHOTODIODE INTO A TYPICAL Si VLSI PROCESS.....	161
6.1 Introduction.....	161
6.2 CMOS Process Sequences	161
6.3 Incorporating the Polysilicon MSM Detector into a Silicon Process	175
Chapter 7 CONCLUSION	179
References.....	185

CHAPTER 1

INTRODUCTION

1.1 Thesis Motivation

Present-day information systems handle increasing data loads. The system throughput of each of these information systems must likewise increase in order to process the ever-increasing volumes of data. This requirement for greater system throughput is placing harsh demands on communication bandwidths as well as processing speeds. In an effort to increase system throughput, designers have made significant progress in increasing processor speeds. Progress in the design of high-speed interconnection networks however, has lagged. The most significant performance bottleneck for the next generation of information systems is the low speed of communications between VLSI chips brought on by the low bandwidth of their interconnection networks [1.1 - 1.5]. In fact, the deciding factor in determining the performance and cost of high-speed computing systems is the communications system between chips and boards which is based on metal interconnections [1.6, 1.7]. Future

demands for greater data rates, interconnection density, and lower error rates are predicted to be well beyond the capabilities of metal interconnections [1.8].

The communications bottleneck imposed by metallic interconnections applies to both uniprocessors and multiprocessor systems. Uniprocessor systems can generally enjoy improved performance with greater clock speeds. As clock speeds increase however, metallic interconnections must be implemented as expensive and bulky high-frequency transmission lines in order to reduce the effects of crosstalk, interference, skin effect, electrical noise, wave reflections, and dielectric imperfections [1.9]. Similarly, multiprocessor systems using metallic interconnections suffer from low communication bandwidth, high power consumption, low interconnection density, and long network latencies [1.10].

High-speed information systems employing metallic interconnections suffer from excessive power consumption and transmission delays, amongst other things. A metallic interconnection is defined as simply any length of conductor connecting two or more circuit nodes. A lumped element electrical interconnection, on the other hand, is a metallic interconnection with the addition of discrete circuit elements such as resistors, capacitors and inductors. A transmission line is defined as a metallic interconnection having a distributed series inductance and shunt capacitance to ground with defined values such that the line has a defined characteristic impedance. In this way a basic metallic interconnection has an undefined impedance whereas a transmission line has a defined impedance value. As illustrated in Figure 1-1, the power required for a lumped element electrical interconnection is proportional to the length of the connection line, whereas the power required for an optical interconnection is practically independent of

connection length for connection less than 1 km in length. While terminated transmission lines have small losses and lose relatively little power per unit length, such transmission lines are costly to build and their power requirement is higher than that for an optical interconnection. As Figure 1-1 shows, electrical interconnections below a certain length require less power than the constant power requirements of optical interconnections and those beyond this length require more power than optical ones. The connection length at which the power requirement curves of optical and electrical interconnections cross is dependent on the technology used in each case and the parts of each system being considered, and different authors claim different crossover points for these curves. One group arrives at a crossover point of 1 cm without considering the power consumed by optical receiver amplifiers or emitter circuitry [1.7]; needless to say, including these circuits would cause the crossover point to occur at a greater connection length. Agreement has been reached, however, that it is most efficient to communicate on-chip electrically, while many off-chip interconnections can be performed most efficiently, from a power consumption viewpoint, by the use of optics. Apart from this power consumption argument, metallic connections require greater chip/board area with increasing interconnection distance since they are confined to pseudo-planar structures. The area required by an two-dimensional interconnection is proportional to its length. The dependence of interconnection power and area requirements on interconnection length drives up the cost of long high-performance metallic interconnections.

The reason why electrical interconnections require more power for longer connection distances stems from a fundamental impedance mismatch existing between the metallic conductors which carry the signals and the logic devices which drive the

Figure 1-1: Relative on-chip power dissipation for electrical and optical interconnections each at 1 Gbit/s transmission rate, showing the connection length at which power dissipation for an optical interconnection equals that for an electrical interconnection.

conductors [1.11]. Owing to the low impedance of free-space (377Ω), and the dependence on conductor separation which dictates that transmission line impedance decreases and capacitance increases as conductor separation decreases, compact electrical transmission lines tend to have low impedance and high distributed capacitance. Electronic drivers, on the other hand, are small logic elements which have large impedance and small capacitance. It is inefficient to use low impedance, high capacitance transmission lines to carry signals driven by high impedance, low capacitance logic devices. The distributed capacitance of a transmission line must first be charged at one end to a logic level and then discharged at the opposite end, requiring both time and power to do so. The sudden change in impedance from the line to the receiver acts as a discontinuity that causes electrical wave reflections and can lead to ringing. One solution which has been adapted is to use large, low-impedance drivers at the transmitting end. These circuits however require greater chip area and more power than smaller drivers;

moreover, they do not remedy the impedance mismatch at the receiver. A separate high-power receiver circuit is often required to solve the mismatch problem at the receiver end.

The power requirement with increasing fan out increases linearly for both electrical and optical interconnections. The energy required to charge a resistive line of capacitance C to a voltage V_o is $E_o = CV_o^2$, where half of this energy is used to charge the line to the required voltage and half is dissipated in the line resistance during charging. This relationship holds even for lossless, properly terminated transmission lines where half of the energy is for the stored electric energy and the other half is for the stored magnetic energy [1.11]. However, these energies are typically much larger in value than the switching energies of small electronic logic devices. To lower the communication energy requires either high impedance lines or lines having low capacitance per unit length, both of which are impractical to make due to size constraints and the relationship between transmission line impedance, capacitance, and lateral dimensions.

Optoelectronic devices, however, perform efficient impedance transformations to match the high impedance of small logic devices to the low impedance associated with signal propagation. A photodetector produces a photocurrent that is proportional to the amount of incident optical power. Optoelectronic devices have a relationship between current and power rather than voltage and power (as is the case for most electrical interconnections). More specifically optoelectronic devices exhibit a relationship between the number of photons generated/absorbed and the number of free electrons produced/required. The minimum amount of optical energy E_o , required to produce a voltage V_o across a photodetector of quantum efficiency β and capacitance C_D is [1.12]

$E_o = h\omega V_o C_D / (\beta e)$ where h is Planck's constant, ω is the photonic angular frequency, and e is the electronic charge. The aforementioned energy is essentially the amount required to generate sufficient electrons to discharge the capacitance of the photodiode in question. Although this is a static result, the dynamic result is not much different as long as $RC_D \geq \tau$, where τ represents the duration of the light pulse and R is load on the photodiode. Notice that the communication energy required for an electrical interconnection is proportional to total line length for all lines in the fan out scheme (or total number of lines where terminated transmission lines are used), whereas that required for an optical interconnection is proportional to the total number of photodiodes in the fan out scheme. In both cases power consumption increases linearly with increasing fan out.

Metallic interconnections have another inherent limitation: the crosstalk which occurs between neighboring conductors. Inductive coupling increases with bit rate and conductor-to-conductor proximity, and so shielding and/or conductor separation is required to avoid crosstalk. Yet another problem with metallic conductors is the timing issue of signals arriving at one point from different source points. Since the time required to charge different metallic lines can vary greatly, the signals may have significantly different propagation times; these timing problems are labeled clock skew and signal skew. The problems of impedance matching, crosstalk, and clock/signal skew are the major factors which limit clock speed.

1.2 Why use Optics?

Optical interconnections offer several advantages over metallic interconnections. They include: higher bandwidth; higher interconnection densities; lower crosstalk which is independent of data rate; inherent parallelism; immunity from electromagnetic interference and ground loops; the ability of free-space optics to fully utilize the third spatial dimension rather than being confined to pseudo-planar structures; lower clock and signal skew; and the capability to perform more fan-ins and fan-outs. While it may not be possible to maximize each of the aforementioned advantages with every optical link, certainly the advantages are numerous enough that the potential benefits of optical interconnections should not be overlooked. Nor will the design of future optical links be any easier than that for electrical interconnections. In fact designers will have significant challenges in creating future optical interconnection systems. As previously stated, the communication bandwidth of metallic interconnections is limited by the capacitance, resistance, and inductance of the link. Optical communications systems however, employ carrier waves, and as such the bandwidth of these systems is limited only by the carrier frequency - about 5×10^{14} Hz for visible light - which is much greater than the 20 GHz mark that electronic signals travelling on terminated transmission lines rarely exceed. Electrical signal propagation speed along metallic lines having undefined impedance depends upon total capacitance, therefore the addition of more fan-outs decreases signal propagation speed due to the added capacitance. Optical signals, on the other hand, propagate independently of the number of devices receiving the signal when proper isolation is used to minimize optical reflections. In this way optical interconnections

allow for a much larger number of fan-ins/fan-outs in addition to the greater transmission speed offered. While optical signals must rely on optoelectronic devices to convert from electrical to optical and back to electrical signals, these devices often use more advanced technologies and are therefore much faster than some of the electrical interconnections that they might replace. For while electronic circuits such as microprocessors can operate at internal clock frequencies of more than 1 GHz, off-chip communications performed with electrical interconnections are more than an order of magnitude lower. Another factor which increases the communication bandwidth available to optical signals is their ability to easily exploit the three spatial dimensions rather than two in the case of electrical signals. The use of the third dimension allows for a much higher interconnection density for optical signals over electrical signals, which gives optical interconnections a potentially higher degree of parallelism.

Optical interconnections perform especially well on the chip-to-chip, board-to-board, and system-to-system levels, though electrical interconnections have a performance to cost ratio that is difficult to beat on the interchip level. On the levels where optical interconnections perform well, they offer additional advantages over conventional metallic interconnections: 1) communicating with optics does not require the electronics to drive a large capacitive load created by the total capacitance of the output bond pad, the signal trace, the bus circuitry if applicable, the input bonding pad, and the input transistor; 2) impedance-matching terminating resistors which increase power dissipation are not required for optical interconnection systems, whereas they are required for longer high speed electrical interconnections; 3) less crosstalk is produced than with metallic connections and; 4) the parallel nature of free-space propagation and

optics' freedom from interference make them ideally suited to perform parallel communications, something highly desirable for parallel processing systems. Regarding the 4th point above, the ability of optical signals to pass through each other without interfering is an advantage when massive parallel interconnections are required. In this way, the interconnections can be made more compact than if the signal lines must maintain a minimal separation distance.

Electrical interconnections are well suited to communications within individual chips. The advantages of metallic interconnections are more prevalent as the interconnection distance shrinks. These interconnections are produced using standard VLSI technology, which is well established, inexpensive and reliable. However, projections of the coming generation of Si CMOS circuits indicate that the density and speed of CMOS based systems will not be limited by the CMOS devices, but by the interconnections between these devices [1.3, 1.4]. Moreover, the Semiconductor Industry Association Roadmap identifies CMOS interconnection technology as a vital technology that must advance in order maintain the historical performance scaling of CMOS technologies. They are unaware of a current technology capable of addressing the interconnection problem beyond the year 2006, but suggest that optical interconnections are potential solutions [1.1].

When comparing optics and electronics, it can be seen that many similarities are shared between the two. Both may be used to carry information, and both rely on electromagnetic waves to do so. Moreover, both may be used to produce the type of nonlinear interaction required to perform logic functions such as switching. However, the frequency of optical waves is much greater than those of electrical waves. Optical waves

of the near infrared spectrum have a frequency of about 3×10^{14} Hz, whereas the frequency of electrical waves seldom exceeds 4×10^{10} Hz, a difference of 4 orders of magnitude. The wavelength of near infrared optical waves is approximately 10^{-6} m, compared to more than 10^{-2} m for electrical waves. That difference means that the construction of waveguides having cross-sectional dimensions that are larger than the wavelength is practical for optical waves and not for electrical waves. The use of dielectric waveguides is possible when the cross-section of the guide is larger than the wavelength of the information carrier; conductors must be used when the wavelength is larger than the cross-section of the guide. Hence it is practical to use dielectric waveguides for optical waves and not for electrical waves. One important difference between the two aforementioned waveguides is in signal loss; dielectric waveguides generally offer substantially less loss than conductive waveguides do.

When performing logic functions using electronics, the required nonlinear effect is obtained through the electrostatic effects of electron interaction. In the case of optical logic, the nonlinear interaction between two light beams is almost always performed by the small movements of electrons inside nonlinear optical materials [1.12]. In this way logic functions are performed by electron interaction both in electrical as well as optical systems, whereas communications is performed by electromagnetic waves in both systems. Therefore, while it is best to perform logic functions (such as switching and routing) with electronics, it is better to use optics for communications over all but very small distances.

Implementation of optical interconnections within an electronic computer holds promise for enhanced computing performance. Electrical interconnections have a

maximum data transfer rate set by the accumulated line capacitance which must be alternately charged and discharged with the signal. In the case of an unmatched transmission line, the energy per bit that is required for communication is proportional to the length of the transmission line, while the bandwidth of the line is inversely proportional to its length. On computer boards, electrical pulses that are lengthy enough to charge the entire length of the interconnections, typically tens of cm, are used to maintain synchronization on a computer board. That a transmission line cannot be produced having an impedance that is known to a sufficiently high degree of accuracy is a consequence of uncertainties in the values of the dielectric constants of the materials used to surround the electrical lines [1.13]. Moreover, for propagation distances greater than 0.1-100 mm required transmitted energy is lower for optical interconnections than in the electrical case [1.11]. Such advantages and the ability of optics to form parallel interconnections have motivated research and development using optics within computers [1.11 - 1.22].

Optical interconnections do have several disadvantages with respect to electrical interconnections, most of which stem from the relative immaturity of optical interconnection technologies. The disadvantages of optical interconnections include the difficulties in aligning optical elements such as transmitters, optical pathways, and detectors, maintaining alignment, as well as their present high cost. The current high-cost of optical links is a result of past optoelectronic research being directed toward long-haul optical links where highest performance was the goal and cost was generally ignored. The cost of optical interconnection bandwidth must decrease to be competitive with that of electrical interconnections before optical links become more pervasive and address the

growing demands for bandwidth Cost reduction of optical links is likely to occur as the technology matures.

1.3 Types of Optical Interconnections

Light makes an excellent information carrier. The ability of a light beam to cross other light beams and emerge unaffected, its high bandwidth, and its immunity from electromagnetic interference combine to make it ideal for use in communications. A pathway must exist, however, for light to travel the route from optical transmitter to receiver. This pathway may take the form of either *free-space* or *guided wave* optics.

Guided wave interconnections require a guiding mechanism for the optical signals. These mechanisms may be either optical fibres or optical waveguides. Guided wave connections, however, when used in the board-to-board or chip-to-chip levels suffer from several problems. The problems associated with fibre-based interconnections include: a relatively low interconnection density; high coupling losses; and high bending losses where the critical angle of optical incidence is exceeded and some fraction of optical power radiates from the fiber core at the bend. High-density interconnection networks requiring many interconnections experience high losses from bending and looping fibres radiating optical power. Problems associated with integrated waveguide interconnections include: coupling of light into and out of the waveguide; coupling-induced crosstalk between nearby waveguides, and the resulting low interconnection densities allowed (this crosstalk becomes more severe in densely connected networks). Moreover, both fibre-based and integrated waveguide interconnections are capable of

performing fewer interconnections and fan-ins/fan-outs than is the case for free-space optics since these interconnections take up physical space, which is limited.

Free-space optical interconnections utilize empty space along with discrete optical elements such as mirrors, lenses, and holograms for optical wave propagation. Free-space interconnections are suitable for board-to-board and chip-to-chip level interconnections because they may potentially exploit the large spatial bandwidth and inherent parallelism of optics. Optics exhibits no mutual interference (light beams can cross without affecting each other as long as the medium is used within its linear limits), and they are able to use simple discrete optical elements which allow high efficiencies and simple architectures. Alignment and stability, however, pose significant problems for free-space optical interconnections.

1.4 Thesis Contributions

This thesis describes the research, design, fabrication, and test of two components of an optical communications link. The components that were produced are a VLSI compatible silicon MSM photodiode, and an integrated silicon microelectronic amplifier designed to amplify the small electrical signal that a photodiode produces. These components are introduced here and described in detail in later chapters.

1.4.1 Silicon-based Photodiode

Optical interconnections are a cost-effective solution for high bandwidth interconnections over optical fibre for computer networking. Two-dimensional guided and free-space optical interconnections have been demonstrated on the backplane level

where the potential aggregate bandwidth exceeds the capability of electrical interconnections. One of the required elements in an optical backplane or interconnect is a low-cost receiver, and the integration of Si-based detectors with Si VLSI circuitry is an active area of current research and development [1.2, 1.6, 1.23].

High-speed integrated photoreceivers have also been demonstrated in GaAs and SiGe [1.24, 1.25]. GaAs ICs have a higher cost and cannot presently be integrated on as large a scale as that achievable with Si. SiGe is also a higher cost technology, but looks promising for use in future photodiodes. Silicon has a higher optical absorption depth ($\sim 10\ \mu\text{m}$ compared with $\sim 0.25\ \mu\text{m}$ for GaAs and $0.3\text{--}10\ \mu\text{m}$ for SiGe at $0.8\ \mu\text{m}$ wavelength), leading to a tradeoff between device speed and responsivity that limits the performance of silicon photodiodes. As metal-semiconductor-metal (MSM) electrode spacing is decreased in an effort to minimize carrier transit-time, the high-field region becomes increasingly shallow [1.26]. More of the carriers are then created in low-field regions and slowly drift toward the MSM contacts, producing a prolonged tail in the current response curve [1.14, 1.15]. It is the relatively long time required for these carriers to either drift to a contact or recombine (the minority carrier lifetime in monocrystalline Si can be as high as $10\ \mu\text{s}$ [1.15]) which makes the fabrication of high-speed Si-based detectors difficult. At short wavelengths ($\lambda = 400\ \text{nm}$), where the absorption depth of Si is only $\sim 100\ \text{nm}$, bandwidths of up to $110\ \text{GHz}$ have been demonstrated [1.27].

In later chapters a novel Si-based MSM detector is described [1.28] which offers good performance characteristics and was fabricated using simple, low-cost, Si VLSI compatible fabrication processes. To the author's knowledge this detector represents the

first use of polysilicon as the photoactive material in a MSM photodiode. As an MSM structure it is a lateral device, and has the advantage of simpler fabrication than *p-i-n* devices, which are vertical structures. Furthermore, polysilicon absorbs strongly in the 800-860 nm range as measurements made and presented in Chapter 3 illustrate, and is therefore compatible with readily available GaAs laser sources. The structure of the detector, which is described in Chapter 6, has the polysilicon layer electrically isolated from its Si substrate by an oxide layer, and ensures that all carriers are generated within a high-field region. The absorption coefficient of polysilicon is higher than that of monocrystalline silicon, and a 2 μm thick layer is sufficient to absorb ~60% of the incident light, thereby ensuring sufficiently high responsivity. Polysilicon is an integral part of standard Si VLSI structures where it is used to form resistors, gate electrodes for MOS devices, and capacitors, and as such is already available for use in nearly all existing Si-VLSI processes.

1.4.2 Amplifier for Optical Receiver

Receivers for optical communication links require an amplifier to transform the weak signal produced by a photodiode to one whose amplitude is compatible with the circuitry which follows the amplifier. Moreover, the receiver must include a low-noise preamplifier so that as little noise as possible is introduced to the weak signal and then amplified by the following amplification stages. An optical link which contains many optical channels must have low per channel power consumption to keep the power consumption of the whole link at a reasonable level, while keeping the noise in each channel at a minimum. It is therefore important to use in this type of link a receiver

amplifier which has sufficiently low power consumption, and at the same time is fast enough to be compatible with the wide bandwidth signals it is designed to amplify.

As part of a collaboration between Carleton University and the National Research Council of Canada (NRC) a monolithic array of 16 receiver amplifiers was designed and fabricated [1.29]. The goal for the amplifier design would be the following properties: a -3 dB point of greater than 400 MHz so that it could work at OC-12 rates, each amplifier would be capable of fully amplifying photocurrents as low as 20 μ A to ECL levels, the amplitude of the output voltage swing be constant and independent of the input signal amplitude, each amplifier would require less than 5 mW of power to operate, and each amplifier was to have an associated output stage which was capable of driving a variety of loads.

One of two different circuit types must be used in order for the amplitude of the output voltage to be independent of the amplitude of the input current signal. One circuit concept uses automatic gain control (AGC), while the other circuit type limits the amplitude of the output voltage signal and as such is called a limiting amplifier. Limiting amplifiers offer many advantages over AGC amplifiers. Limiting amplifiers generally require less power for operation than AGC amplifiers do; they have simpler designs than AGC amplifiers, and therefore require less chip area; and they generally have no need for external circuitry, as opposed to AGC amplifiers.

A circuit principle proposed by Cherry and Hooper [1.30] was employed to give the amplifier the widest possible bandwidth. This principle involves the arrangement of transimpedance and transadmittance amplifier stages placed alternately in a series configuration and, as discussed in chapter 5, is often used in wideband amplifier design.

A transimpedance stage is characterized by the presence of voltage-shunt feedback, as well as low input and output impedances. A transadmittance stage, on the other hand, is characterized by the presence of current-series feedback and high input and output impedances. The mismatched impedances of the alternating transimpedance and transadmittance stages allows each stage to operate with very little loading effect from the following stage at up to very high frequencies. The feedback becomes less effective as circuit parasitics begin to dominate, causing the impedance mismatch between the stages to degrade at higher frequencies, along with the amplifier performance.

Given the desired low level of power consumption, the amplifier had to be a single-ended rather than differential design, thereby foregoing the advantages of higher speed, higher dynamic range, better noise immunity and reduced crosstalk provided by differential stages. One of the few advantages single-ended design has over differential is lower power consumption. Moreover, the amplifier was designed to use a 3.0 V supply, and the entire circuit was modeled numerous times to minimize the current passing through each transistor in order to minimize the power consumption of the circuit. Given that single-ended circuits are more susceptible to crosstalk, more care must be taken to ensure that isolation between stages is adequate.

1.5 Thesis Outline

This thesis describes the work that went into the two project contributions: the polysilicon MSM photodiode and a silicon integrated preamplifier/limiting amplifier. Chapter 1 is a thesis introduction. Chapter 2 provides an overview of optical detection

and the methods of transforming an optical signal into an electrical one. Chapter 3 concentrates on metal-semiconductor-metal (MSM) diodes as well as the optical and electrical properties of polycrystalline silicon. Chapter 4 explains the design and measurements made on the polysilicon MSM photodiode, while Chapter 5 details the design and measured results of the silicon integrated preamplifier/limiting amplifier circuit. Chapter 6 describes how a typical modern CMOS process could be modified to allow for the integration of the polysilicon MSM photodiode onto a CMOS wafer. The final chapter, Chapter 7, is a summary chapter that provides a synopsis of the thesis contributions.

CHAPTER 2

OPTICAL DETECTION

2.1 Photons in Semiconductors

This chapter includes some brief discussion on semiconductor physics that will be useful in the discussion on applications. It is assumed that the reader has had some exposure to a treatment of solid state physics at the level of the book by Kittel [2.1]. In the interests of brevity, some topics are not fully discussed here, although references to more complete treatments are given.

2.1.1 The Concept of the Photon

The concept of the photon was initially proposed by Planck in 1900 and then refined by Einstein in 1905. Planck found that he could explain the phenomena of blackbody radiation by assuming that electromagnetic energy was emitted and absorbed in discrete amounts. Later, after studying the photoelectric effect, Einstein concluded that visible light itself was distributed in discrete “quanta” of energy. Up until the time at

which the idea of the photon was put forth, the best theories on light stated that it was a continuous wave, however we now know that light, and indeed all electromagnetic energy, is quantized into small packets called photons. The energy of a photon was determined by Planck to be directly proportional to its frequency:

$$E_p = h\nu = hc / \lambda \quad (2-1)$$

where

$h \cong 6.626 \times 10^{-34}$ J-s is Planck's constant;

ν is the frequency of electromagnetic radiation;

$c \cong 2.998 \times 10^8$ m/s is the speed of light in a vacuum;

λ is the wavelength of the radiation.

Electromagnetic radiation, or light, however can at times also appear to behave as a wave. As the flux of photons through a given area increases, the intensity of the radiation field increases and wave properties become more evident. Suffice it to say that although light is comprised of discrete particles, it can also exhibit the characteristics of a wave, including all of the wave properties such as reflection, refraction, diffraction, and interference. Electromagnetic radiation is therefore classified by its photon energy E_p and by its intensity.

2.1.2 Atomic Energy Levels

The Danish physicist Niels Bohr postulated in 1913 that electrons existed in certain distinct orbits (called stationary states) around the nucleus of an atom, and that

electrons could move freely within these orbits without radiating energy. He then assumed that the atom radiated when one of its electrons made a transition from one stationary state to another and that the frequency of the radiated energy is related to the energies of the stationary states by

$$h\nu = E_i - E_f \quad (2-2)$$

where E_i and E_f are the total energies of the initial and final states.

Atoms, molecules, and crystal lattices have specific allowed energy states which are determined by the laws of quantum mechanics. When isolated from each other, atoms and molecules have discrete energy levels, as illustrated in Fig. 2-1. A photon may donate its energy to an atom by raising it to a higher energy state if the photon's energy matches the energy difference between the two levels. In this way the photon is said to be absorbed by the atom. In a reverse process, the atom can emit a photon by falling to a lower energy state.

Figure 2-1: Electron energy level of hydrogen, showing the n discrete energy levels.

Solids, on the other hand, are composed of atoms and molecules which are spaced close together and cannot be treated as simple isolated particles, but rather as many-body

systems. Whereas energy levels in isolated atoms and molecules are discrete, those in the crystal lattice of a solid are broadened to take the form of energy *bands*. The band structure of a crystalline solid describes the relationship between energy and momentum ($E-k$) of electrons in the solid, where the solution of the Schrodinger equation of an approximate single-electron problem [2.2]

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_k(\mathbf{r}) \right] \phi_k(\mathbf{r}) = E_k \phi_k(\mathbf{r}) \quad (2-3)$$

gives band structure, where E_k is the energy of the electron, m is the mass of the electron, and V_k is the potential field confining the electron. The Bloch theorem states that the eigenfunctions of the wave equation for a periodic potential are the product of a plane wave $\exp(j\mathbf{k} \cdot \mathbf{r})$ times a function $U_k(\mathbf{r})$ having the periodicity of the crystal lattice. The solutions $\phi_k(\mathbf{r})$ of the Schrodinger equation to band structures are called Bloch functions:

$$\phi_k(\mathbf{r}) = e^{j\mathbf{k} \cdot \mathbf{r}} U_k(\mathbf{r}) \quad (2-4)$$

where $U_k(\mathbf{r})$ has the period of the crystal lattice with $U_k(\mathbf{r}) = U_k(\mathbf{r} + \mathbf{T})$, $\mathbf{T}$ being the period of the lattice. From the Bloch theorem, it can be shown that the wavevector $\mathbf{k}$ is sufficient to determine the energy of the band.

In a crystal containing N primitive unit cells, each primitive cell contributes one independent value of the electron wavevector $\mathbf{k}$, and since there are two possible orientations of electron spin (up and down), there exist $2N$ independent states in each energy band of the band structure [2.1]. Therefore, each energy band can hold $2N$ electrons. For example, consider a lattice in which the primitive cells contain single atoms of valence one, the band would be half filled with electrons. If the cells contain a

single atom of valence two each, or if they contain two atoms of valence one each, then the band would be exactly filled.

If the valence electrons exactly fill some of the energy bands, leaving the others completely empty, the material will be an insulator. This situation is illustrated in Fig. 2-2, which shows the highest filled valence band and the lowest empty band along with the energy gap between the two bands. An externally applied electric field will not cause a current to flow through an insulator because there is no continuous way to change the total electron momentum of the lattice when all the accessible states are occupied, provided that the filled band and the next higher empty band are separated by an energy gap.

A crystal can be considered an insulator only if the number of valence electrons in its primitive cell is an even integer. Satisfying this requirement, it is still necessary to determine whether there exists an energy gap between successive bands. An energy band contains as many discrete energy levels as the number of atoms which contribute to the formation of the band. Since only two electrons can occupy the same level, due to electron spin considerations and the Pauli exclusion principle, an energy band may need

Figure 2-2: Energy band diagram of an insulator or semiconductor showing the highest filled valence band and lowest empty band separated by an energy gap E_g .

twice as many electrons as there are atoms in the material to fill the band. If successive bands overlap and no energy gap exists between them, then instead of having one filled band, there are two half filled bands and the crystal is considered a metal. A crystal is also considered to be metallic if it contains an odd number of valence electrons per primitive cell. A metal is any material that possesses only partially filled energy bands. Figure 2-3 illustrates schematically the band structure of a metal. A semiconductor differs only quantitatively from an insulator - the energy gap between bands in a semiconductor is smaller than that of an insulator (~ 1 eV as compared to several eV). Due to the small energy gap or band gap between energy bands in the crystal's electron band structure, a few electrons will have sufficient energy to be thermally excited into the lowest empty energy band at temperatures above 0 K, the quantity of electrons excited into the next band being dependent on the temperature. The addition of these few electrons to the previously empty band means that they are now mobile electrons, giving the crystal a weak electrical conductivity. Because of this thermally created conductivity, the lowest empty band at 0 K is called the conduction band

Figure 2-3: Schematic illustration of the energy band structure of a metal.

Energy bands existing in solids can often be explained by the nearly free electron model [2.1] for which the band electrons are assumed to be weakly perturbed by the periodic potential of the positive ion cores of the lattice. Consider a linear solid of lattice constant a . There will be a gap in this solid's band structure caused by Bragg reflection of electron waves by the crystal lattice; because for Bragg reflection, wavelike solutions of the Schrodinger equation do not exist [2.1]. Figures 2-4(a) and (b), respectively, illustrate qualitatively part of the band structure for a free electron, and a nearly free electron which has an energy gap at $k=\pm\pi/a$. The Bragg condition $(\mathbf{k}+\mathbf{G})^2=k^2$ for the diffraction of a wave having wavevector $\mathbf{k}$ becomes in one dimension $k = \pm G/2 = \pm n\pi/a$, where $G = 2\pi n/a$ is a reciprocal lattice vector and n is an integer. The Bragg condition means that the first reflection and energy gap occurs at $k = \pm\pi/a$, and subsequent reflections and energy gaps occur for other values of the integer n .

The wavefunctions at $k = \pm\pi/a$ are not the travelling waves $\psi_{\mathbf{k}}(\mathbf{r}) = \exp(j\mathbf{k}\cdot\mathbf{r})$ of free electrons but instead are composed of equal parts of waves travelling to the right and to the left, and as such they are standing waves, waves which do not move with time. We can form two different standing waves from the two travelling waves $\exp(\pm j\pi x/a)$:

$$\begin{aligned}\psi(+) &= \exp(j\pi x/a) + \exp(-j\pi x/a) = 2\cos(\pi x/a); \\ \psi(-) &= \exp(j\pi x/a) - \exp(-j\pi x/a) = 2j\sin(\pi x/a)\end{aligned}\tag{2-5}$$

The probability density ρ of a particle is $\psi^*\psi = |\psi|^2$. A travelling wave $\exp(jkx)$ has a probability density $\rho = \exp(-jkx)\exp(jkx) = 1$, indicating that the electron distribution and charge density is uniform. However, for standing waves the charge density is not constant. The standing wave $\psi(-)$ of (2-5), for example, gives the probability density $\rho(-) = |\psi(-)|^2 \propto \sin^2(\pi x/a)$ which is highest between lattice sites and therefore concentrates

Figure 2-4: Plots of energy E vs. wavevector k for (a) a free electron and (b) an electron in a linear lattice of lattice constant a , showing an energy gap at $k = \pm\pi/a$.

electrons between ions of the lattice. On the other hand, the standing wave $\psi(+)$ gives a probability density $\rho(+) = |\psi(+)|^2 \propto \cos^2(\pi x/a)$ which concentrates electrons on the ions.

By calculating the expected potential energy over the three charge distributions discussed, it can be found that the potential energy of $\rho(-)$ is higher than that of the travelling wave, whereas the potential energy of $\rho(+)$ is lower than that of the travelling wave. The difference between the energies of $\rho(+)$ and $\rho(-)$ is E_g and that is the width of the energy gap of the material, as shown in Fig 2-4. Figure 2-4 also shows point A where the wavefunction is $\psi(-)$ and point B, just below A, where the wavefunction is $\psi(+)$.

2.1.3 Occupation of Energy Levels

Atoms and molecules continuously undergo random transitions between their different energy levels. Each of the molecules of a gas which is in thermal equilibrium at temperature T has a probability P of being in or above a particular energy state E_m as given by the Boltzmann distribution

$$P(E_m) \propto \exp(-(E_m - E_o)/kT), \quad m = 1, 2, 3, \dots \quad (2-6)$$

where k is the Boltzmann constant and E_o is the average energy of the molecules. The Boltzmann distribution gives an occupation probability which is an exponentially decreasing function of E_m .

In a gas, it is possible for many molecules to possess the same energy value. In the case of electrons however, only two can occupy the same energy level, though they must be of opposite magnetic spin. Electrons are spin-1/2 particles (*fermions*), and as such the energy levels of electrons are split into positive and negative magnetic spin states. Moreover, the Pauli exclusion principle tells us that each state can be occupied by at most one electron. This restriction on the number of particles occupying the same energy level alters the energy distribution slightly from Equation (2-6). At temperature 0 K all electrons in a solid occupy the lowest possible energy level. For a semiconductor, this means that at 0 K the valence band is completely filled and the conduction band is completely empty. However, as the temperature is raised the number of electrons possessing sufficient thermal energy to escape to the conduction band likewise increases. The probability that an electron occupies a given energy level E of a solid in thermal equilibrium is given by the *Fermi-Dirac distribution* [2.1]:

$$f(E) = \frac{1}{\exp\left[(E - E_f)/kT\right] + 1} \quad (2-7)$$

where E_f is a constant called the *Fermi energy* or *Fermi level*. The energy level E is either occupied, with probability $f(E)$, or unoccupied, with probability $1-f(E)$. An unoccupied energy state in the valence band is called a *hole*. Zero and integer spin particles (*bosons*) such as α particles, photons, and most atoms and molecules (most atoms and molecules have zero spin) obey a slightly different distribution law [2.3].

2.1.4 Absorption of Photons in Semiconductors

The atoms of a semiconductor are capable of emitting and absorbing photons, just as other atoms are, by transferring energy between a photon and an electron state. In the case of semiconductors, the change in electron energy corresponds to the difference in energy levels of the valence and conduction bands or between dopant levels within the energy gap. This means that direct band to band absorption (no transitions through intermediate states such as traps) can occur only for photon energies $h\nu > E_g$, the photon wavelength which this corresponds to is called the bandgap wavelength λ_g . If a photon has an energy greater than or equal to E_g it may excite an electron from the valence band to the conduction band, thus creating an electron-hole pair. Additional electron-hole pairs increase the concentration of mobile charge carriers and increase the conductivity of a material. In this way the semiconductor behaves as a photoconductor whose conductivity is proportional to the flux of photons incident upon it.

A transfer of energy is not always all that is necessary for the absorption and emission of photons: some semiconductors require a change in electron momentum as well. Consider a free electron, its energy E and momentum $\mathbf{p}$ are related by $E = p^2/2m_0 = \hbar^2 k^2/2m_0$, where p is the magnitude of the momentum and k is the magnitude of the wavevector $\mathbf{k} = \mathbf{p}/\hbar$ of the electron's wavefunction, and m_0 is the mass of the electron. The electron energy E is proportional to the square of the wavevector k and so the graph of the E - k function is a simple parabola. The electrons and holes of a semiconductor however behave differently, their motions being determined by the Schrodinger equation and as such their E - k relations are not simple parabolas.

The E - k relations for semiconductors contain a set of solutions for the conduction as well as the valence bands, and at energies near their band minima and maxima, each may be approximated by parabolas. Semiconductors whose conduction band energy minima correspond to the same wave vector k as the energy maxima of the valence band are called *direct-gap* materials. Semiconductors whose energy minima and maxima of their conduction and valence bands, respectively, do not correspond in momentum are called *indirect-gap* materials; Fig. 2-5 illustrates an example of each. Silicon is an example of an indirect-gap semiconductor, while GaAs is a direct-gap semiconductor. The difference in band structure of the two types of semiconductor means that a change in electron momentum is required for a band transition in an indirect-gap material, whereas no momentum change is required for a direct-gap material. Momentum must always be conserved in the process of photon absorption. This means that the momentum of the electron produced must equal the momentum of the hole produced minus the negligible

Figure 2-5: Approximations of the E - k relations near the top of the valence band and the bottom of the conduction band for (a) Si, an indirect-gap semiconductor and (b) GaAs, a direct-gap semiconductor.

(for photons within the visible and infrared spectrums) momentum of the photon involved.

Photon absorption is less likely in indirect-gap semiconductors than it is in direct-gap semiconductors [2.4]. An electron in the valence band most likely resides near the top of that band, and the conservation of momentum rule requires that, without a change in momentum, the electron moves vertically in the band diagrams of Fig. 2-5. For a direct-gap semiconductor the transition to the bottom of the conduction band (where electrons are most likely to reside in this band) is a single step process, since the bottom of the conduction band has the same momentum as the top of the valence band. The excitation of an electron to the bottom of the conduction band in an indirect-gap semiconductor however, is a two-step process. The minimum energy gap of the band structure involves electrons and holes having substantially different wavevectors $\mathbf{k}_c$, and as such a direct transition to the bottom of the conduction band cannot satisfy the conservation of momentum rule. But if a crystal vibration or phonon of wavevector $\mathbf{K} = -\mathbf{k}_c$ and frequency Ω is created in the process then momentum will be conserved. The phonon energy $\hbar\Omega$ means that the minimum photon energy for absorption is now $\hbar\omega = E_g + \hbar\Omega$, where $\hbar\Omega$ is generally much less than E_g . For the aforementioned reasons, direct-gap semiconductors are more efficient at absorbing photons than indirect-gap semiconductors and hence produce better photodetectors.

2.2 Properties of Semiconductor Photodetectors

This section is devoted to semiconductor photodetection. Common fundamental rules determine the operation of all semiconductor photodetectors such as the photoconductor, photodiode, and avalanche photodiode. All three of these types rely on the photoexcitation of electrons from the valence to the conduction band to convert optical energy to an electrical current. All three also share the basic photodetector properties of quantum efficiency, responsivity, and response time.

2.2.1 Quantum Efficiency

A photodetector's quantum efficiency is defined as the probability that a given photon incident on the device has of generating an electron-hole pair which contributes to detector current. In other words, quantum efficiency is the ratio of electron-hole pair generating photons to total photons incident on the detector. Since it is a probability, the quantum efficiency η is a fraction which is greater than zero but less than one ($0 \leq \eta \leq 1$). The whole absorption process is a probabilistic one: a photon passing an atom has a probability of giving up its energy to that atom, and therefore the more atoms it passes the higher the probability that a photon has of being absorbed. The equation which describes the attenuation of an optical wave travelling through some medium is $\exp(-\alpha d)$, where d is the distance through the medium which the light travels, and α is a constant called the *absorption coefficient*. The fraction of light absorbed is simply one minus the above attenuation factor or $1 - \exp(-\alpha d)$. Some photons, however, are reflected at the surface and do not enter the photosensitive region of the detector; and some electron-hole pairs

produced near the detector surface recombine and do not contribute to detector current.

The quantum efficiency η of a detector can be expressed as

$$\eta = (1 - R)\zeta[1 - \exp(-\alpha d)] \quad (2-8)$$

where R is the reflection coefficient of the detector surface and represents the fraction of incident optical power that is reflected off of the detector surface, and ζ is the fraction of electron-hole pairs which contribute to detector current. Some definitions of quantum efficiency exclude reflectance at the detector surface, which is considered separately in such cases.

2.2.2 Responsivity

The ratio of current flowing in the detector to incident optical power is called the detector's *responsivity*; it relates the incident photon flux to the electron flux in the detector, and is therefore related to quantum efficiency. A photon flux Φ having an optical power $P = h\nu\Phi$ incident onto a detector produces an electron flux, and gives an electrical current $i_p = \eta e\Phi$ in the device. The two expressions for power and current can be combined to give the electrical current in terms of incident optical power.

Some detectors do not produce a charge e for each photogenerated carrier, but exhibit an internal gain; such detectors create a charge q for each carrier. The gain G is defined to be the average number of current-contributing electrons produced per photon generated electron-hole pair, and is distinguished from η , the probability that an incident photon has of creating a detectable electron-hole pair. The gain $G = q/e$ may be either greater or less than unity. Where G is greater than unity, the total number of charge carriers generated exceeds the number of photogenerated charge carriers. Where G is less

than unity, the total number of charge carriers is less than the number of original charge carriers generated by photoexcitation. The photocurrent generated by such detectors is

$$i_p = \eta q \Phi = G \eta e \Phi = \frac{G \eta e P}{h \nu} = \mathfrak{R} P \quad (2-9)$$

where $\mathfrak{R}$ is the constant of proportionality, having units A/W, which relates photocurrent to optical power and is called responsivity:

$$\mathfrak{R} = \frac{G \eta e}{h \nu} = \frac{G \eta e \lambda}{h c} \quad (2-10)$$

where c is the speed of light in a vacuum. $\mathfrak{R}$ is directly proportional to λ (where the constant of proportionality is $G \eta e/h \nu$) as seen in Equation (2-10), because it is the flux of photons rather than the incident optical power that the detector responds to. The spectral range over which responsivity increases with wavelength λ is limited, since photons with λ high enough that their energies are less than the energy gap of the semiconductor of the detector cannot generate an electron-hole pair in the detector, and so do not contribute to detector current.

2.2.3 Response Time

The charge delivered to the external circuit of a photodetector by charge carrier motion does not appear instantaneously, but must first travel to the edge of the semiconducting material. This effect is called *transit time spread*, and it limits the speed of response of all semiconductor photodetectors. An electron-hole pair created by the absorption of a photon in a semiconductor will ideally drift toward opposing device contacts and ultimately cause one electron to leave the negative terminal and one electron to enter the positive terminal (holes do not exist outside the semiconductor material).

Thus, each charge-carrier pair created causes one electron to flow in the external current. However, because each charge-carrier travels at a finite velocity, each contributes some current to the external circuit for the entire time it takes each carrier to reach its respective terminal. The effect of transit time spread is to distribute the flow of electrical charge over a finite time span.

The impulse response of a photodetector is an important parameter in determining the speed of response in the time domain. The impulse response is the photocurrent $i_p(t)$ that flows in the external circuit due to an incident optical impulse. It is determined by the number of photogenerated carriers produced by the optical impulse, the locations at which these carriers are created, and the speeds at which they move to their respective terminals. The current due to electron flow of N_e free electrons in a detector of width w is

$$i_e(t) = \frac{N_e e}{w} v_e \quad (2-11)$$

where v_e is the average electron velocity. A similar equation for holes is

$$i_h(t) = \frac{N_h e}{w} v_h \quad (2-12)$$

The total photocurrent is the sum of $i_e(t)$ and $i_h(t)$

$$i_p = \frac{e}{w} [N_e(t)v_e + N_h(t)v_h] \quad (2-13)$$

Both N_e and N_h decrease with time since electrons and holes continuously leave the photoactive region after an optical impulse is incident on the detector.

Consider a simple case where the density of electron-hole pair generation is uniform throughout the photoactive region of the detector and the charge carriers move

with constant velocity, v_e and v_h . In such a case the initial values of N_e and N_h are equal, $N_e(0) = N_h(0) = N_o$ and the initial photocurrent is:

$$i_p(0) = \frac{eN_o}{w}(v_e + v_h) \quad (2-14)$$

As electrons are swept out of the photoactive region with velocity v_e , N_e decreases

$$N_e(t) = N_o \left(1 - \frac{t}{\tau_e}\right) \quad (0 < t < \tau_e) \quad (2-15)$$

where $\tau_e = w/v_e$ is the full device width electron transit time. This gives

$$i_e(t) = \frac{eN_o v_e}{w} \left(1 - \frac{t}{\tau_e}\right) \quad (2-16)$$

and similarly,

$$i_h(t) = \frac{eN_o v_h}{w} \left(1 - \frac{t}{\tau_h}\right) \quad (2-17)$$

Since v_e is generally greater than v_h , the initial amplitude of i_e is greater than that of i_h as shown in Figure 2-6. The i_e pulse is also shown to have a shorter duration than that of i_h .

The total impulse response of the detector is the sum of $i_e(t)$ and $i_h(t)$, and is also illustrated in Figure 2-6.

Another factor that limits the response time of semiconductor detectors is the RC time constant created by the resistance R and the capacitance C of the detector and its circuitry. Usually, one of these two effects will dominate and the speed of response of the detector can be well estimated by considering the dominant effect solely. However, if the two effects are of the same order of magnitude, then the impulse response in the presence of transit-time and RC time constant effects can be determined by convolving $i(t)$ from transit-time considerations with the exponential function $(1/RC)\exp(-t/RC)$ [2.4].

Figure 2-6: Photodiode impulse response for uniform generation of electron-hole pairs throughout detector active region

2.2.4 Photoconductors

As mentioned previously, when a semiconductor absorbs a photon, mobile charge carriers (an electron-hole pair) are created. In this way the electrical conductivity of the semiconductor increases with the incident photon flux. By applying an electric field to the material by way of an external voltage source these free charge carriers can be made to drift, causing current to flow. Photoconductor detectors operate by registering either the photocurrent directly, or the voltage drop across a load resistor R in series with the detector. Photoconductors are made by attaching an anode and a cathode each to a piece of semiconductor by way of simple ohmic contacts (a contact that obeys Ohm's law, unlike a diode contact). The contacts are often placed on the same face of the semiconductor material.

The change in photocurrent is caused by a change in incident photon flux Φ illuminating a semiconductor volume wA [2.4]. A fraction η of the incident photons are

absorbed to create excess electron-hole pairs. The production rate R of carrier pairs per unit volume is given by $R = \eta\Phi/wA$, where w is the width of the photoactive region and A is its cross-sectional area. The rate of excess carrier recombination is given by $\Delta n/\tau$ where Δn is the photoelectron concentration and τ is the carrier recombination lifetime. Under steady state conditions the rates of electron production and recombination are equal ($R = \Delta n/\tau$), giving $\Delta n = \eta\tau\Phi/wA$. The increase in excess carrier concentration gives rise to an increase in electrical conductivity expressed as

$$\Delta\sigma = e\Delta n(\mu_e + \mu_h) = \frac{e\eta\tau(\mu_e + \mu_h)}{wA}\Phi \quad (2-18)$$

where μ_e and μ_h are the electron and hole mobility values. The increase in conductivity, as illustrated in Equation (2-18), is therefore proportional to the incident flux of photons. Current density J_p is related to conductivity as $J_p = \sigma E$ and the carrier velocities are $v_e = \mu_e E$ and $v_h = \mu_h E$, where E in this case represents the electric field; using these expressions and Equation (2-18) it can be shown that $J_p = [e\eta\tau(v_e + v_h)/wA]\Phi$. The electric current flowing is then $i_p = AJ_p = [e\eta\tau(v_e + v_h)/w]\Phi$. If $v_e \gg v_h$ and $\tau_e = w/v_e$, the current may be expressed as

$$i_p \approx e\eta\frac{\tau}{\tau_e}\Phi \quad (2-19)$$

The ratio τ/τ_e in Equation (2-19) above determines the detector gain $G = \tau/\tau_e$.

Assuming that electrons travel faster than holes and that the recombination lifetime τ is much longer than the electron transit time, electrons travelling to one end of the semiconductor under the influence of the applied electric field will reach their destination before the holes will. As one electron leaves the semiconductor another one must be

supplied by the external circuit at the opposite end of the detector in order to maintain charge neutrality within the device. The new electron then drifts toward the opposite end of the detector; this process is repeated until an electron finally recombines with a hole or the hole completes its trip and enters the external circuit. In this way it is possible for several electrons to enter the external circuit and contribute to detector current for each photon absorbed, thus producing gain.

The response time of photoconductors is usually constrained by transit-time effects since they generally have very small capacitance and hence a small RC time constant. Their carrier-transport response time is approximately equal to their recombination time τ . This means that the detector's carrier-transport bandwidth B is inversely proportional to τ . Since the detector's gain G is directly proportional to τ , there exists a trade-off between gain and bandwidth and the gain-bandwidth product GB is roughly independent of τ . Given that the response time of a detector is dominated by transit-time effects, GB is determined by carrier velocity, or τ_e to be more precise.

2.2.5 Photodiodes

A photodiode is simply a p - n junction for which the reverse current increases when photons are absorbed. Photodiodes rely on photogenerated charge carriers to produce an electric signal from an optical one just as photoconductors do. Unlike photoconductors however, p - n and p - i - n photodiodes can be made to produce no gain. It is possible however to make p - n and p - i - n photodiodes that produce significant gain by supplying a bias voltage that is large enough for avalanche multiplication to occur.

A reverse-biased p - n junction diode under illumination absorbs photons throughout its bulk, regardless of their direction of incidence, with an absorption coefficient α . The absorption of photons in the diode generates electron-hole pairs, but only in the diode's depletion layer where there exists an electric field will a significant fraction of the total charge carriers generated be transported out of the diode and into the external circuit. Since the electric field in a diode points from the n -region to the p -region, electrons drift toward the n side and holes move toward the p side. Electron-hole pairs generated far away from the depletion layer are rarely transported because there is no electric field outside of the depletion layer and there is little chance that a charge carrier will diffuse into the depletion layer. However, charge carriers generated near the depletion layer have a chance of entering the depletion region by the random effects of diffusion, the rate of diffusion being affected by temperature and the spacing between atoms of the crystal through which the diffusion occurs. Once inside the depletion region, the effects of the electric field present there sweep the charge carriers across it.

The response time of a photodiode is determined by transit time effects and its RC time constant, as are other photodetectors, but photodiodes also must contend with the effects of carrier diffusion. Those carriers generated outside of the diode's depletion layer must diffuse into it in order to contribute to the detector current, and carrier diffusion is a much slower process than carrier drift under the effects of an electric field. Photodiodes are still generally faster than photoconductors because of the large electric field that exists within the photodiode's depletion region and produces the much higher carrier velocities.

Photodiodes are generally operated in strong reverse bias; this biasing has many advantages. A strong reverse bias produces a large electric field within the depletion region of the diode; this in turn imparts a higher drift velocity to the charge carriers and reduces carrier transit time. A strong reverse bias also increases the depletion layer width, and therefore decreases the junction capacitance as well as the RC time constant, giving a faster response. Another benefit of the wider depletion layer is the greater volume of the depletion region, thus allowing the absorption of more photons and giving the photodiode a higher quantum efficiency.

A $p-i-n$ photodiode is a $p-n$ junction diode with a lightly doped, or intrinsic, layer sandwiched between the p and n layers. The advantage that the $p-i-n$ diode has over the $p-n$ diode is its much wider depletion layer. The depletion region can be made effectively as wide as the intrinsic layer of the diode. The wider depletion region increases the diode's quantum efficiency by increasing the charge-collecting volume of the photodiode. It also means a lower junction capacitance and hence a lower response time. Too wide an intrinsic layer, however, increases the diode's response time due to increasing transit times. Another advantage of increasing the depletion width is that it reduces the ratio between the diffusion length and the drift length of the detector. This results in a greater fraction of the photogenerated current being carried by carrier drift rather than the slower diffusion process.

Schottky-barrier photodiodes, another type of photodiode, are formed from metal-semiconductor heterojunctions. Metallic electrodes are used in place of a p or n -type layer in a $p-n$ junction photodiode. The Schottky-barrier structure and its energy-band diagram are illustrated schematically in Figure 2-7. The height of the potential barrier

Figure 2-7: Energy bands for a metal and an n -type semiconductor (a) separate and (b) bonded together. The potential barrier is shown in heavy lines.

formed by the Schottky contact is $W - \chi$, where W is the work function of the metal, defined as the energy difference between the metal's Fermi level and the ionization energy or vacuum level, and χ is the electron affinity measured from the bottom of the conduction band to the vacuum level. The vacuum level is the potential energy of a free electron relative to the nearest ion core, or the energy required to free an electron from a metal.

An avalanche photodiode (APD) is another type of photodiode that should be discussed due to their popularity amongst optical link designers. APDs produce gain by converting each detected photon into a deluge of mobile charge carrier pairs. Such an avalanche of carriers is produced by strongly reverse-biasing a photodiode and thereby creating a large electric field within the device. The electric field must be great enough to accelerate charge carriers to speeds which give them sufficient energy E_g to create other carrier pairs in the semiconductor through a process called *impact ionization*. The absorption of a photon in the depletion region of an APD typically produces the following chain of events. The absorbed photon creates an electron-hole pair which move in

opposite directions under the influence of the strong electric field. The accelerating charge carriers occasionally collide with the crystal lattice, losing some of their energy in the process. Should either of the carriers acquire an energy greater than the band gap of the semiconductor E_g between collisions, it will have the opportunity to generate another electron-hole pair at the next lattice collision. The newly generated carrier pair similarly separates and its two parts accelerate in opposite directions, parallel to the electric field, and may act as the source for further impact ionization. As this process continues, an avalanche of mobile charge carrier pairs is created, hence the name avalanche photodiode.

There has been much work lately on photodiodes for use with silicon VLSI technology. The best choice of wavelength for short-haul optical interconnections is 850 nm due to the low cost of emitters and detectors operating in this vicinity. While the optical losses at 850 nm are not as low as for some of the longer wavelengths, attenuation incurred in short-distance communication is not significant, and the cost effectiveness of other parts of the communications link become more important. For this reason it is desirable to design and fabricate a photodetector that is compatible with silicon VLSI technology. Metal semiconductor metal photodiodes (see Chapter 3) and *p-i-n* photodiodes are currently attracting the greatest attention from researchers. The performance of several of these recently reported photodiodes at approximately 850 nm is summarized in Table 2-1. The goal of all of the photodiodes presented in Table 2-1 is a good combination of high bandwidth, high efficiency, and ease of integration with silicon VLSI technologies. As Table 2-1 shows, the bandwidths of these recently published photodiodes range from several hundred megahertz to several gigahertz, while their responsivities range from 0.01 to 0.57 A/W. These performance parameters are

significantly less than those of commercially available photodiodes, which can have values of the same performance parameters that are more than an order of magnitude greater. For short-haul applications though, highest performance is not the only system goal; system cost is also an important metric.

Table 2-1: Comparison of several recently published photodiodes for use with 850 nm

Technology	P.D. type	Dark Current (nA)	P. D. Area (μm^2)	Cap. (fF)	Responsivity (A/W)	-3dB Bandwidth (GHz)	Ref.
Silicon	MSM	30 000	10^4	–	0.40	2.05	[2.5]
Silicon	MSM	20	10^4	–	0.25	2.05	[2.5]
Silicon	<i>p-i-n</i>	–	2700	–	0.30	0.55	[2.6]
Silicon	<i>p-i-n</i>	–	274	–	0.01 – 0.04	1.0	[2.7]
Silicon	<i>p-i-n</i>	2.1	2346	–	0.20	0.90	[2.8]
Silicon	<i>p-i-n</i>	0.116	2346	–	0.57	0.124	[2.8]
SiGe	MSM	1000	–	–	<0.12	0.93	[2.9]
SiGe	<i>p-i-n</i>	1000	156	–	0.3	0.45	[2.10]
GaAs	MSM	–	9×10^4	2×10^3	0.26	1.0	[2.11]
GaAs	MSM	5	1963	68	0.45	5.9	[2.12]

CHAPTER 3

METAL-SEMICONDUCTOR-METAL PHOTODIODES

3.1 Theory of the MSM Diode

The metal-semiconductor-metal photodiode (MSM-PD) is a two terminal device in which a uniformly doped semiconductor lies between two metal contacts, as illustrated in Figure 3-1(a). The energy band diagram for such a structure, which is equivalent to two Schottky diodes placed back to back, is given in Figure 3-1(b). ϕ_{n1} and ϕ_{n2} here are the barrier heights for the two contacts, V_{D1} and V_{D2} are the built-in potentials for contacts 1 and 2 respectively, ϕ_{p2} is the hole barrier height for contact 2, and w is the thickness of the semiconductor between the electrodes. In the case of equivalent contact areas in the MSM diode and similar contact metals, thereby equating the two barrier heights, the structure is called a symmetrical MSM diode. By altering the doping level in the semiconductor of a MSM diode, one can change the energy band structure of the diode. Note, however, that although the doping levels change, the potential step heights ϕ_n and ϕ_p remain constant, as illustrated in Figure 3-2.

Figure 3-1: Schematic view of (a) a metal-semiconductor-metal structure, and (b) the corresponding energy band diagram with no applied voltage across the MSM diode. ϕ_{n1} and ϕ_{n2} are the electron barrier heights, ϕ_{p2} is the hole barrier height for contact 2, and V_{D1} and V_{D2} are the built-in potentials for contacts 1 and 2 respectively.

By applying an external voltage to the MSM structure, the sum of the two depletion widths of the two back to back diodes increases. As the applied voltage increases, eventually the two depletion regions will meet and the composite depletion region will have a width equal to w , the thickness of the semiconductor. The voltage at which this joining of depletion regions occurs is called the reach-through voltage V_{RT} . Reach-through is illustrated in Figure 3-3(a). Further increases in the applied voltage shift the flat part of the energy band, and the zero electric field point, all the way to the position $x = w$; this situation is called the flat-band condition and it is illustrated in Figure 3-3(b).

Figure 3-2: Energy band diagrams for MSM structures having semiconductors which are (a) n - type, (b) p - type and (c) intrinsic semiconductor.

Figure 3-3: Energy bands and electric fields of MSM structures at (a) reach through where the two depletion regions meet each other and the gradient of the electric field E is constant so that the semiconductor is completely depleted of free charge, (b) the flat band condition for which the energy band at position $x = w$ becomes flat, and (c) the condition for an applied voltage V larger than the flat-band voltage V_{FB} .

The flat band condition requires that the flat-band voltage V_{FB} be applied to the external circuit. Applied voltages greater than V_{FB} cause the slope of energy band to grow even steeper, while the slope of the electric field curve remains constant, as illustrated in Figure 3-3(c). The maximum applied voltage is limited by avalanche breakdown in the semiconductor near contact 1, where the electric field is greatest as Figure 3-3 shows.

In order to respond well to absorbed photons, the MSM diode should be biased at a voltage which is greater than the flat band voltage V_{FB} . At applied voltages less than V_{FB} a potential well is formed, as shown in Figure 3-3(a), which traps electrons. Flat band voltage is related to electrode spacing w and the effective carrier concentration N_D for a symmetrical MSM structure by [3.1]:

$$V_{FB} = \frac{eN_D}{2\epsilon_s} w^2 \quad (3-1)$$

where e is the electronic charge and ϵ_s represents the permittivity of the semiconductor. As stated previously, the maximum V_{FB} is the voltage at which avalanche breakdown occurs near the reverse biased contact.

The barrier heights ϕ_{n1} and ϕ_{p2} for the biased MSM structure are not the same as those for the unbiased structure of Figure 3-1. The Schottky effect causes an image-force induced lowering of the potential barriers when an electric field is applied [3.1]. The Schottky barrier lowering $\Delta\phi$ is given by:

$$\Delta\phi_{n1} = \sqrt{\frac{eE_{m1}}{4\pi\epsilon_s}} \quad (3-2)$$

where E_{m1} is the magnitude of the electric field at contact 1, and for an MSM structure biased at voltage V beyond the reach-through condition, E_{m1} is given by:

$$E_{m1} = \sqrt{\frac{2eN_D}{\epsilon_s} V} \quad (3-3)$$

The Schottky effect and the image-force lowering of potential barriers are illustrated in Figure 3-4.

The barrier height of ϕ_{n1} determines the amount of electron current at the reverse-biased contact, since there is an exponential relationship between electron injection and ϕ_{n1} . Similarly, the hole injection at the forward-biased contact is determined exponentially by the barrier height $\phi_{p2} = E_g - \phi_{n1}$. The assumption that $\phi_{p2} + \phi_{n1} = E_g$ is valid when both the anode and cathode are formed from the same metal deposition process. Although a small contribution to device dark current is due to the spontaneous, thermally controlled creation of electron-hole pairs within the active region of the semiconductor, the bulk of dark current is due to carriers which overcome the Schottky

Figure 3-4: The potential profile of an MSM structure under bias showing the applied voltage V and the image force barrier lowerings $\Delta\phi_{n1}$ and $\Delta\phi_{p2}$, as well as the directions of electron and hole motion under the applied electric field.

barrier at the device contacts. Total dark current density for an MSM structure biased higher than the reach-through voltage is therefore approximated by [3.1]:

$$J_{n1} = A_n^* T^2 e^{-\frac{e}{kT}(\phi_{p2} - \Delta\phi_{p2})} + A_p^* T^2 e^{-\frac{e}{kT}(\phi_{n1} - \Delta\phi_{n1})} \quad (3-4)$$

where A_n^* and A_p^* are the effective Richardson constants for electrons and holes respectively, and T is the absolute temperature. According to Equation (3-4), there exists a trade-off between electron current and hole current in that as the electron barrier is lowered at the expense of the hole barrier, less electron current and more hole current will result. It is necessary to choose the proper metal for contacts 1 and 2 to give the correct balance between ϕ_{p2} and ϕ_{n1} , and to thereby provide the lowest dark current. It is generally desirable to use a metal at the anode and cathode which has a barrier height close to $E_g/2$; this will usually provide the lowest dark current. Table 3-1 [2.2] lists the measured Schottky barrier heights of various metals bonded to n -type Si which are close

Table 3-1: Measured Schottky barrier heights at 300 K.

Metal	Schottky barrier height (V)
Cu	0.58
Al	0.72
Ti	0.50
Ni	0.61
W	0.67

to half the bandgap of Si itself, for which $E_g = 1.12$ eV. Ti was chosen as the Schottky metal based on its Schottky barrier height which is close to $0.5E_g$, as well as its excellent adhesion to silicon.

To understand how a metal-semiconductor contact creates a Schottky barrier, consider two limiting cases. The first case involves an ideal metal-semiconductor contact which is free of surface states. Consider again the energy band diagrams of a

Figure 3-5: Energy levels for a typical metal and semiconductor. The metal is described by a work function $e\Phi$, whereas the semiconductor is described by an electron affinity $e\chi$ as well as a band gap ($E_c - E_v$).

metal and a semiconductor as illustrated in Figure 3-5. The uppermost energy level is the vacuum or free-electron energy E_o , it acts as a reference level and represents the energy needed to free an electron of the influence of a given material. Another energy level represented is the Fermi level. The Fermi level is the energy level at which the probability of an energy state being filled by an electron is equal to 50%. At 0 K, the Fermi level is the highest filled energy level.

The difference between the vacuum level E_o and the Fermi level E_f is called the work function (symbolized as $e\Phi$ in energy units or just as Φ in volts). In metals, the work function is a constant but in semiconductors, where the Fermi level is a function of doping, the work function is not a constant. The difference between the vacuum level and the conduction band edge, however, is a material constant of a semiconductor. This energy difference is called the electron affinity, and is denoted $e\chi$ in energy units.

If the metal and semiconductor of Figure 3-5 were brought into contact, electrons from the conduction band of the semiconductor would flow into the metal since Φ_M is greater than Φ_S , and hence the average energy of electrons in the semiconductor is greater than that of the metal. Electrons will continue to flow from one material to the other until the Fermi levels in the two materials become equal. Now a band diagram of the metal-semiconductor contact can be created; the sketch is given in Figure 3-6. Note that the vacuum level E_o is drawn as a continuous curve, this is because it represents the energy required to free an electron from the material which holds it, and as such must change as the properties of a material change. Note also that electron affinity remains constant, since it is a constant of a given material.

As Figure 3-6 illustrates, a sharp discontinuity is formed in the allowed energy states at the metal-semiconductor junction, the magnitude of this barrier $e\phi_B$ is given by:

$$e\phi_B = e\Phi_M - e\chi \quad (3-5)$$

As depth into the semiconductor increases, the energies of electrons at both band edges changes depending on doping, work functions and electron affinities. The change in electron energy is due to the transfer of charge across the interface. The charge transfer creates an electric field near the junction, and since the carrier concentration in the semiconductor is much lower than that of the metal, the electric field reaches a much greater depth in the semiconductor than it does in the metal. The effect of the field within the semiconductor is to deplete the free-electron population near the junction. The depletion of free-electrons is indicated in Figure 3-6 by the increased separation between E_c and E_f near the junction in comparison to that of the bulk semiconductor.

Several assumptions have been made about the semiconductor charge in Figure 3-6. The charge that is transferred from the semiconductor to the metal is assumed to exist only on the metal surface. In the n-type semiconductor of Figure 3-6, positive charge can exist in the form of ionized donors or holes in the valance band. It is assumed that the hole population is everywhere negligible. It is also assumed that the electron concentration is much less than the donor concentration in a region (called the space-charge region) which extends from the semiconductor surface to a distance x_d into the semiconductor, since donor concentration is equal to electron concentration at depths greater than x_d from the semiconductor surface. While these approximations are not precisely correct, they are sufficient to give an understanding of the nature of the metal-

Figure 3-6: Idealized equilibrium in a metal-semiconductor Schottky contact: a) energy band diagram; b) charge distribution at the junction, note that the negative charge is approximately a delta function at the metal surface; c) Electric field at the junction.

semiconductor interface. The maximum electric field ξ_{max} is located at the interface and is given by Gauss' law as [2.2]:

$$\xi_{max} = \frac{-eN_d x_d}{\epsilon_s} \quad (3-6)$$

where ϵ_s is the permittivity of the semiconductor. The potential difference across the space-charge region ϕ_i is equivalent to the negative of the area under the electric field curve and is given as:

$$\phi_i = -\frac{1}{2} \xi_{max} x_d = \frac{eN_d x_d^2}{2\epsilon_s} \quad (3-7)$$

By solving for x_d in Equation (3-7) the amount of space-charge $Q_s \equiv eN_d x_d$ in the semiconductor may then be written as:

$$Q_s = eN_d x_d = \sqrt{2e\epsilon_s N_d \phi_i} \quad (3-8)$$

The second limiting case occurs when the metal and semiconductor are separated by a thin interfacial layer which is characterized by the presence of surface states. This

interfacial layer is too thin to keep electrons from tunneling through it, however it can sustain a voltage drop. Figure 3-7 illustrates the condition of a metal-semiconductor contact with the presence of a thin interfacial layer sandwiched between the metal and semiconductor. A concentration of D_s states exists on the surface bounding the semiconductor and interfacial layer; these states are assumed to be of the acceptor type, varied in their energy, and independent of the metal. The interfacial layer is of thickness δ and sustains an energy drop $q\Delta$ in the vacuum level, and since this layer is thin enough to allow electron tunneling through it, the electron-semiconductor barrier height is measured from the Fermi level to the conduction band at the semiconductor surface. Even in the absence of a metal, the interfacial layer is present and charge is transferred from the semiconductor in order to balance the charge present in the surface states. This charge transfer effectively adds charge to some of the surface states and places those states below the Fermi level. D_s and the position of the valance band edge govern the amount of charge removed from the semiconductor as follows: as D_s increases, so does

Figure 3-7: Band structure of a metal-semiconductor interface with the presence of surface states. The model incorporates a thin interfacial layer of thickness δ which sustains a voltage drop Δ at equilibrium. Surface states are assumed to be acceptor-type and are present in concentration D_s .

the amount of charge removed for every incremental increase in the valance band edge.

Therefore if D_s is sufficiently large, a negligible movement of the Fermi level at the semiconductor surface (caused by the presence of a metal, for example) would cause sufficient charge to flow *from the surface* to equalize the Fermi levels. A *pinning* of the Fermi level at a constant level above the semiconductor's valence band is the result.

While the Fermi level in the semiconductor does not move with respect to the valence or conduction band, it does move to align itself with the Fermi level of a metal that is brought into contact. This phenomenon is possible because the vacuum level in the semiconductor shifts with respect to that of the metal due to the voltage drop Δ across the interfacial layer. Note that both acceptor and donor surface states can cause pinning of the Fermi level when the state density becomes sufficiently large. A more comprehensive analysis of metal-semiconductor contacts with surface states has been presented [3.2] which goes into more detail than that which has just been presented.

When the density of surface states is large enough to pin the Fermi level, the barrier height $e\phi_B$ is defined by:

$$e\phi_B = (E_g - e\phi_o) \quad (3-9)$$

where $e\phi_o$ is the difference between the Fermi level and valance band edge at a semiconductor surface which is not in contact with a metal. As the surface state density D_s approaches zero, the barrier height $e\phi_B$ approaches that predicted by basic Schottky theory (Equation (3-5)). The value of D_s determines whether a metal-semiconductor barrier height corresponds more closely to that predicted by basic Schottky theory or Equation (3-9). In practice, most metal-semiconductor barrier heights are more

accurately predicted by Equation (3-9) than by basic Schottky theory. It is therefore the surface properties which determine the barrier height most often, rather than the work function of the metal. This is true for silicon, germanium, and gallium arsenide.

The contacts of an MSM photodiode usually take the form of an interdigitated structure, in this way maximizing the photosensitive area while minimizing the distance between electrodes. Figure 3-8 illustrates schematically the electrode layout of a typical MSM photodiode.

The capacitance of an interdigitated MSM photodiode can be calculated using conformal mapping techniques [3.3]. The total capacitance is given as:

$$C = C_o(N-1)l \quad (3-10)$$

where N is the number of fingers in the MSM diode structure, l is the finger length, and C_o is the gap capacitance of electrodes per unit length and is given by:

$$C_o = \varepsilon_o(1 + \varepsilon_r)K/K' \quad (3-11)$$

where ε_o is the dielectric constant of free space, ε_r is the relative dielectric constant of the semiconductor, and K and K' are the complete elliptic integrals defined as follows:

$$K = K(k) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}} \quad (3-12)$$

$$k = \tan^2 \left(\frac{\pi}{4} \cdot \frac{W}{W + L} \right)$$

$$K' = K(k'), \quad k' = \sqrt{1 - k^2} \quad (3-13)$$

where W is the width of the electrode fingers and L is the width of the space between fingers (see Figure 3-8). An approximation for K/K' given by Hilberg [3.4] for the case where $0 < k < 1/\sqrt{2}$ is given by:

Figure 3-8: Schematic view of a MSM photodetector.

$$\frac{K}{K'} = \frac{\pi}{2} \ln \left\{ 2 \sqrt{\frac{1+k'}{1-k'}} \right\} \quad (3-14)$$

This approximation is valid for K/K' over the range: $0 < K/K' < 1$. The case where $0 < k < 1/\sqrt{2}$ represents interfinger spacing in the range $0 < L < 8W$, which encompasses most, if not all, MSM interdigitated photodiode structures which typically exhibit a width to spacing ratio W/L on the order of unity. The capacitance of typical MSM photodiodes is much less than that of a conventional $p-i-n$ photodiode of equal photosensitive area, however the usable photosensitive area of an interdigitated MSM detector is a fraction, approximately $L/(W+L)$, of that of a similarly sized $p-i-n$ detector.

The response speed of an MSM photodiode can be found by taking the Fourier transform of the impulse response of the device. Since the MSM photodiode has a very small capacitance, the RC time constant of the circuit will typically be much less than the photodiode's impulse response time and therefore can be neglected when calculating response time. Several assumptions are made for the purposes of simplicity: 1) the carriers move with constant drift velocity within the photosensitive region; 2) light is absorbed uniformly within the photosensitive region, producing an initially uniform

carrier density and; the electrode walls are parallel and facing each other. These assumptions produce a scenario, when an optical impulse is absorbed, in which n_o electron-hole pairs per unit volume are created. Under these conditions, the total number of charge-carriers will decrease linearly with time as they drift towards, and are collected by, their respective electrodes. In this way, the impulse current density will be the sum of the separate electron and hole impulse current densities, which are:

$$\begin{aligned} J_e(t) &= qn_o \frac{w}{\tau_e} \left(1 - \frac{t}{\tau_e}\right) & 0 \leq t < \tau_e \\ &= 0 & t \geq \tau_e \end{aligned} \quad (3-15)$$

$$\begin{aligned} J_h(t) &= qn_o \frac{w}{\tau_h} \left(1 - \frac{t}{\tau_h}\right) & 0 \leq t < \tau_h \\ &= 0 & t \geq \tau_h \end{aligned} \quad (3-16)$$

where

$$J_{imp}(t) = J_e(t) + J_h(t) \quad (3-17)$$

Figure 3-9 illustrates graphically the normalized impulse response current density.

Taking the Fourier transform of the impulse response gives the frequency response of the MSM photodiode:

$$J(f) = \int_{-\infty}^{\infty} J_{imp}(t) \exp(-i2\pi ft) dt \quad (3-18)$$

which can be simplified to give:

$$J(f) = \frac{-i}{2\pi f \tau_e} + \frac{-i}{2\pi f \tau_h} + (2\pi f)^{-2} (\tau_e^{-2} + \tau_h^{-2}) - \frac{\exp(-i2\pi f \tau_e)}{(2\pi f \tau_e)^2} - \frac{\exp(-i2\pi f \tau_h)}{(2\pi f \tau_h)^2} \quad (3-19)$$

Figure 3-9: Normalized impulse response current density of a MSM photodiode.

The frequency response of the current density $J(f)$, when $\tau_e = 300$ ps and $\tau_h = 600$ ps, is graphed in Figure 3-10. The -3dB point on the graph is at $f = 1.12$ GHz.

The responsivity of a MSM photodiode can be calculated using Equations (2-8) and (2-10). A semiconductor photodetector with 100% quantum efficiency detecting light of wavelength 800 nm has a responsivity $\mathfrak{R} = 0.64$ A/W. When the thickness of semiconductor is not infinite but rather a more realistic value of $1/\alpha$, the responsivity is $\mathfrak{R} = 0.40$ A/W. Taking into account the reflectivity of the top surface of a detector (0.30 in the case of an air-silicon interface) the responsivity drops to $\mathfrak{R} = 0.28$ A/W. Finally, in

Figure 3-10: Calculated normalized frequency response of polysilicon MSM photodiode having a 2 μm thick photosensitive layer. The -3 dB point is at $f = 1.12$ GHz.

the case of an interdigitated MSM photodiode where metallic electrodes block some fraction of the incident light, the responsivity drops to $\mathfrak{R} = 0.14$ A/W when the electrode fingers and intrafinger spaces are of equal width and the above limitations are also considered.

3.2 Photoconducting Materials

The choice of photoconducting material for an MSM photodiode is an important one. The material has a great effect on the response speed of the device as well as the ease of integration with electronic circuitry. The material also determines the wavelength range of the photons the detector is capable of detecting. Direct bandgap materials are sought because of their high absorption coefficients. Materials with high carrier mobility values are sought because of the high response speed they impart onto the detectors they are fabricated from.

The materials of choice for discrete, high-speed, high quantum efficiency photodetectors are the III-V compound materials. Gallium arsenide (GaAs) for example has an absorption coefficient of $2 \times 10^4 \text{ cm}^{-1}$ at 300 K for 800 nm wavelength photons. The high absorption of GaAs allows for the production of high quantum efficiency detectors, while simultaneously it has electron and hole mobility values of $8500 \text{ cm}^2/\text{V-s}$ and $400 \text{ cm}^2/\text{V-s}$ respectively, which make it a good material to use in the fabrication of fast response-time detectors. The economics of long-distance communications dictate that

1.3 μm or 1.5 μm wavelength radiation be used, and for these wavelengths some of the III-V compounds are most effective.

Crystalline silicon, on the other hand, has electron and hole mobility values of 1500 $\text{cm}^2/\text{V-s}$ and 450 $\text{cm}^2/\text{V-s}$, respectively, and its absorption coefficient at 800 nm and 300 K is only $1.5 \times 10^3 \text{ cm}^{-1}$. It is fairly obvious that crystalline silicon does not perform as well as GaAs in the areas mentioned above. Silicon does, however, perform very well in the area of integration with VLSI circuitry, which is built almost exclusively on silicon, and while GaAs MSM photodiodes require a thin passivation layer on top of the bulk GaAs to lower their dark current, Si MSM photodiodes require no passivation layer [3.5]. Although some III-V semiconductors work well for 1.3 μm and 1.5 μm wavelengths, these wavelengths are not required for short-distance communications. The losses incurred in short-distance communication are not significant, and the cost effectiveness of other parts of the communications link become more important. From this point of view, 0.8 μm becomes the best choice of wavelength due to the low cost of emitters and detectors operating in this vicinity. Silicon becomes a very good detector at 0.8 μm . For the preceding reasons it is desirable to design and fabricate a photodetector having a silicon photosensing layer.

3.2.1 Amorphous Silicon

Alternatives to crystalline silicon exist which have a much higher absorption coefficient than crystalline Si and at the same time possess reasonably high mobility values. Amorphous silicon has been measured to have an absorption coefficient of 4×10^4

cm^{-1} at 300 K for 600 nm wavelength radiation, while having an electron mobility higher than $1.0 \text{ cm}^2/\text{V-s}$ [3.6]. Damage introduced by ion implantation has also been shown to increase the absorption coefficient of Si, while at the same time lending it a much greater electron mobility ($20 - 100 \text{ cm}^2/\text{V-s}$) than that of amorphous silicon [3.7, 3.8].

Increased light absorption and lower carrier mobility values in amorphous and damaged Si as compared to crystalline Si is a result of the non-crystalline and disturbed crystalline structure in the aforementioned materials. The Bloch theorem, mentioned earlier in section 2.1.2 which explains the broad semi-continuous energy bands found in crystalline solids, does not apply where the material's structure is not periodic. Because every atomic site in the solid may have different distances to each neighboring site, electron states cannot be described by well-defined $\mathbf{k}$ values. In this scenario, the momentum difference between valence band peak and conduction band minimum is not well-defined, and the transition from one to the other is able to proceed without a required change in the electron's momentum. In this way, photons are absorbed much more readily than in a perfectly ordered crystalline structure. Generally speaking, any interruption of the crystalline order which exists in solid silicon will act to increase the optical absorption of the sample.

The decrease in carrier mobility is caused by the same perturbations in crystal structure. The carriers simply cannot move as fast in a disordered structure as they can in a uniform crystal lattice, due to the abundance of defect sites found in the former structure type which provide band-tail and gap states. The transport of excess carriers in a disordered structure involves trapping, detrapping, and transport in extended states [3.9]. Carrier transport in extended states and detrapping are both slow processes. Crystalline

defects are also responsible for shorter carrier lifetimes since the defect sites act as efficient recombination centres for charge carriers. Conduction in polysilicon is discussed in section 3.2.3.

3.2.2 Optical Absorption in Polycrystalline Silicon

Polycrystalline silicon exhibits disorder at the boundaries between crystal grains. For this reason it has enhanced optical absorption, lower carrier mobility values and shorter carrier lifetimes than monocrystalline silicon - just as is the case for amorphous silicon, and for the same reasons. Various types of photoconductors have been demonstrated which use polysilicon as the absorbing material [3.10, 3.11], and a polysilicon *p-i-n* photodiode has also been demonstrated [3.12]. A trade-off exists between optical absorption and carrier lifetime together with carrier mobility. The grain size in the polycrystalline Si to a large extent determines the tradeoff point, with a larger grain size producing higher mobility and a small grain producing higher optical absorption. One study [3.13] found that polycrystalline silicon films deposited by low-pressure chemical vapour deposition at 625 °C (so that they were composed of columnar grains and similar to those films used in IC fabrication) *had an optical absorption coefficient α for 780 nm wavelength radiation of $\alpha = 5 \times 10^3 \text{ cm}^{-1}$* . Tests performed by the author (of this thesis) on similarly-grown films of polycrystalline silicon gave comparable results, see Figure 3-11. This optical absorption coefficient represents a substantial improvement over that of monocrystalline silicon. It gives polycrystalline silicon an absorption depth $1/\alpha$ of 2 μm compared to 10 μm for monocrystalline silicon.

In order to determine the optical absorption coefficient of polycrystalline silicon, the author performed measurements of a sample. The sample was produced in the Carleton University, Department of Electronics' fabrication lab by low-pressure chemical vapour deposition at 625 °C by thermal decomposition of silane gas (SiH₄). The polycrystalline silicon was grown to a measured thickness of 2.4 µm on one side of a 2 inch diameter fused silica wafer. A Perkin-Elmer Lambda 9 spectrometer was then used to measure the absorbance of the sample for a range of wavelengths of incident light. The absorbance A is defined as:

$$A = \log_{10} \left(\frac{I_o}{I_T} \right) \quad (3-20)$$

where I_o is the incident optical intensity, and I_T is the transmitted optical intensity. From the absorbance data, the absorption coefficient can be determined. When the sample is a single layer of a substance, the relationship between transmitted and incident optical intensity can be found by taking the infinite sum of the waves reflected inside the sample, just as is the case for a Fabry-Perot etalon [3.14]. The zeroth wave passes through the sample with no reflections, the first wave passes with a single round trip reflection inside the sample, the second wave passes with two round trip reflections inside the sample, and so on for s total waves. This sum of reflected waves represents a simple geometric series:

$$\sum_{n \rightarrow \infty} r^{n-1} = \frac{1}{1-r} \quad \text{if } |r| < 1 \quad (3-21)$$

and the infinite sum of waves through the sample is approximated accordingly as:

$$I_r = \frac{I_o(1 - R_1)(1 - R_2)}{\exp(\alpha d) - 2\sqrt{R_1 R_2} \cos(4\pi n d / \lambda)} \quad (3-22)$$

where R_1 is the reflection coefficient of one side of the sample, R_2 is the reflection coefficient of the other side of the sample, α is the absorption coefficient of the sample material, n is the refractive index of the sample, λ is the wavelength of the light passing through the sample, and d is the distance through the sample material which the incident light travels. Rearranging the above Equation (3-21), the absorbance can be expressed as:

$$A = \log_{10} \left(\frac{I_o}{I_r} \right) = \log_{10} \left(\frac{\exp(\alpha d) - 2\sqrt{R_1 R_2} \cos(4\pi n d / \lambda)}{(1 - R_1)(1 - R_2)} \right) \quad (3-23)$$

Rearranging the above equation, the absorption coefficient can be expressed as:

$$\alpha = \frac{\ln \left[(1 - R_1)(1 - R_2) 10^A - 2\sqrt{R_1 R_2} \cos(4\pi n d / \lambda) \right]}{d} \quad (3-24)$$

The samples that were tested consisted of two layers: 2.5 μm of polysilicon grown on top of a 600 μm thick quartz wafer. The Perkin-Elmer was calibrated with a bare quartz wafer identical in construction to the one on which the polysilicon layer was grown such that the absorbance of the quartz wafer was calibrated out of the samples.

The reflection coefficient at normal incidence R of an interface can be determined from the index of refraction of the two sides of the interface, n_1 and n_2 respectively:

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2 \quad (3-25)$$

The reflection coefficient R_1 is that of the air-polysilicon interface, and R_2 is that of the poly-quartz interface. The refractive index of polysilicon n_p is $n_p = 3.42$; the refractive index of quartz n_q is $n_q = 1.55$; and the refractive index of air n_a is $n_a = 1.00$. Absorption

coefficient data collected from the polycrystalline silicon sample are plotted against incident photon wavelength in Figure 3-11.

To investigate the optical absorption of large-grained polysilicon, a sample was produced by low-pressure chemical vapour deposition of silane by thermal decomposition at a temperature of 560 °C. This deposition produced a layer of amorphous silicon 0.6 μm thick. The sample was then annealed at 550 °C in a N_2 atmosphere for 88 hours. This anneal was performed to produce polysilicon with larger grain size than that of the previous sample. This sample was also measured in the Perkin-Elmer Lambda 9. The measured optical absorption coefficient, however, was close to that of monocrystalline silicon, and thus provides little advantage over monocrystalline silicon when intended for use in an optical detector.

Figure 3-11: The measured absorption coefficient of our polysilicon samples. The bandgap energy of polysilicon is 1.1 eV and is therefore transparent to photons having less than this energy, this includes light of wavelength 1.3 μm and 1.55 μm . It does however have a fairly high absorption coefficient for shorter wavelength light such as that for which $\lambda = 800\text{nm}$, for which the photon energy is 1.55 eV.

3.2.3 Electrical Properties of Polycrystalline Silicon

Polycrystalline silicon is generally deposited by chemical vapour deposition at temperatures of approximately 625°C, involving the thermal decomposition of silane [3.15]. A literature search for measured electron mobility in this type of polycrystalline silicon film -the type used most often in IC fabrication - has uncovered little. One study has discovered hole mobility values as high as 45 cm²/V·s [3.16]. These samples, although deposited at 650 °C, were subsequently annealed at 1100 °C, thereby increasing the grain size beyond that obtained from a 650 °C deposition only. As such, data from them cannot be considered valid for such a deposition. Under normal deposition conditions, grain structure is columnar with column diameters in the range 20 - 100 nm. The columnar structure of polysilicon gives it anisotropic electrical conduction properties. Effective carrier mobility is generally higher in the direction parallel to the long axis of the columnar grains [3.17]. Thin films of polysilicon generally will have narrower grains than those of thick films; grains start out in narrow columns which increase in diameter with distance from substrate surface.

Conduction in polycrystalline silicon is via a different process than that which occurs in monocrystalline silicon. A polycrystalline material is composed of small crystallites joined together by grain boundaries. The grain boundary usually consists of a few atomic layers of disordered atoms, whereas the interior of the crystallites is an ordered arrangement of atoms that may be considered as a small single crystal. The grain boundary acts as a transitional region between two crystallites which can have vastly different angles of crystallographic orientation. The presence of grain boundaries in

polycrystalline silicon causes the conduction process in it to differ from that of monocrystalline silicon.

The most widely accepted theory of electrical conduction within polycrystalline silicon involves carrier trapping at the grain boundary [3.16-3.23]. The theory reasons that since much disorder exists between atoms of the grain boundary, there is a large number of defects which produce incomplete atomic bonds, or dangling bonds. These incomplete bonds act as efficient carrier traps and result in the formation of trapping states. These trapping states are capable of trapping and immobilizing carriers, after which the traps become electrically charged. An electrically charged trap creates a potential barrier which impedes the motion of a carrier from one crystallite to the next. In this way the presence of grain boundaries, and therefore traps, retards the motion of carriers and therefore reduces the mobility of these carriers.

Dopant atoms in polycrystalline silicon contribute free carriers to the conduction or valence band of the semiconductor. Many of these free carriers however, are trapped at low-energy positions created by the dangling bonds located at grain boundaries and are unable to contribute to conduction. The trapped charge within grain boundaries is balanced by an equal amount of opposite charge, in the form of depletion regions, on either side of the grain boundary (see Figure 3-12). Charge in depletion regions cause their energy bands to curve, according to Poisson's equation, and therefore leads to the formation of potential barriers in the energy bands of polycrystalline silicon [3.16-3.23]. It is these potential barriers that retard the motion of carriers from one grain of silicon to another.

Figure 3-12: (a) A one-dimensional model for the structure of polysilicon. (b) Charge distribution within the crystallite and at the grain boundary. (c) The energy band structure of polysilicon showing potential barriers at grain boundaries.

Using Poisson's equation, the barrier height V_B (Figure 3.12c) in a one-dimensional case can be expressed in terms of the dopant concentration N :

$$\frac{d^2 V}{dx^2} = \frac{qN}{\epsilon_s} \quad (3-26)$$

where x is the spatial dimension, e is the charge on an electron, and ϵ_s is the permittivity of silicon. By solving for V_B , the above Equation (3-26) can be written as:

$$V_B = \frac{qN}{2\epsilon_s} x_d^2 \quad (3-27)$$

where x_d is the width of the depletion region on either side of the grain boundary.

As the above Equation (3-27) shows, the barrier height depends on the dopant concentration - it also depends on the trap density and energy. At low dopant concentration the total number of carriers NL per unit area of a length L of polysilicon is much less than the number of traps N_T per unit area found at the grain boundary. Given that the energy levels of the traps are sufficiently low, nearly all of the carriers provided

by donor atoms will be trapped at the grain boundaries. If the dopant concentration is low enough, the depletion regions surrounding the grain boundaries widen to include the entire grain; and the curvature of the energy bands, as illustrated in Figure 3-12, is slight. Low curvature of energy bands means that the potential barrier located at the grain boundaries are small, and that carriers are able to readily move from one grain to another.

As the concentration of dopant rises, grain boundary traps collect more carriers, the energy bands curve more, and the potential barriers associated with grain boundaries become higher. It is assumed in the carrier-trapping model that the trap energy is low enough that the traps are completely filled when the dopant concentration is greater than the critical value $N^* = N_T/L$ [3.16]. As such, when the dopant concentration exceeds N^* , the number of carriers trapped in states at the grain boundary remains constant and the excess carriers form neutral regions within the crystallites. The neutral regions centred around the core of each of the grains causes the width of the depletion regions to shrink to a value x_d :

$$x_d = \frac{N_T}{2N} \quad (3-28)$$

and the barrier height can, in this case, be written as:

$$V_B = \frac{qN}{2\epsilon_s} \left(\frac{N_T}{2N} \right)^2 = \frac{qN_T^2}{8\epsilon_s N} \quad (3-29)$$

so that as the dopant concentration increases beyond $N = N^* = N_T/L$ the potential barrier decreases. It can be seen that the addition of dopant atoms to polycrystalline silicon has the effect of increasing the depletion region barrier height until the concentration reaches

the critical value N^* , at which point the barrier height then begins to decrease with increasing dopant concentration. Fig. 3-13 illustrates this point.

Figure 3-13: The potential barrier associated with grain boundary depletion regions increases in height linearly with dopant concentration N up to a critical point $N = N_T/L$, after which it becomes inversely proportional to N .

The width of the depletion region on either side of the grain boundary is small enough only at very high dopant concentrations to allow for carrier tunneling through the depletion region potential barrier. Most often the carriers rely on thermionic emission over the barrier to travel from one grain to the next. When thermionic emission dominates, the current density J is dependent on an applied electric field as well as the height of the depletion region potential barrier. With zero applied field, the potential barrier is of equal height in both forward and reverse directions, and the net current flow is zero. In the presence of an applied field however, the potential barrier decreases in one direction and increases in the opposite direction; this differential in barrier height causes a net current to flow in one direction. It has been shown [3.21] that, in general, the applied voltage is not divided equally between the two sides of a grain boundary, but for small

applied fields the voltage V_G across a single grain boundary is split approximately equally between each side of the boundary. Thus the current density in the forward direction is:

$$J_F = qnv_c \exp\left[-\frac{q}{kT}\left(V_B - \frac{V_G}{2}\right)\right] \quad (3-30)$$

and the current density in the reverse direction is:

$$J_R = qnv_c \exp\left[-\frac{q}{kT}\left(V_B + \frac{V_G}{2}\right)\right] \quad (3-31)$$

where n is the free carrier density, V_B is the potential barrier height without an applied field, V_G is the bias applied across the depletion region, and v_c is the collection velocity defined as:

$$v_c = \sqrt{\frac{kT}{2\pi m^*}} \quad (3-32)$$

where m^* is the effective mass of the carrier. The net current density is then:

$$\begin{aligned} J_T &= qnv_c \exp\left(-\frac{qV_B}{kT}\right) \left[\exp\left(\frac{qV_G}{2kT}\right) - \exp\left(-\frac{qV_G}{2kT}\right) \right] \\ &= 2qnv_c \exp\left(-\frac{qV_B}{kT}\right) \sinh\left(\frac{qV_G}{2kT}\right) \end{aligned} \quad (3-33)$$

It has been reported [3.23] that conduction in polycrystalline silicon agrees with this expression. At low applied bias the voltage drop across each grain boundary is small when compared to the thermal voltage kT/q and the following approximation may be used:

$$\sinh\left(\frac{qV_G}{2kT}\right) \approx \frac{qV_G}{2kT} \quad (3-34)$$

The net current density can then be given by:

$$J_T \approx \frac{q^2 n v_c}{kT} V_G \exp\left(-\frac{qV_B}{kT}\right) \quad (3-35)$$

In order to express the current density in the standard form of $J_T = qn\mu E$ an effective low-field mobility μ_{eff} can be defined:

$$\mu_{eff} = \frac{qv_c L}{kT} \exp\left(-\frac{qV_B}{kT}\right) \quad (3-36)$$

This is not the same as the usual mobility, which is related to isolated scattering centres in a monocrystalline piece of semiconductor. Instead it describes the obstruction of current flow by potential barriers located at grain boundaries. As the effective mobility is dependent exponentially on barrier height, it is strongly dependent on dopant concentration. As stated previously, barrier height increases linearly with increasing dopant concentration up to a critical concentration N^* , and as such, effective mobility decreases with increasing dopant concentration up until the same point is reached. Beyond the concentration N^* , the potential barrier height decreases and the effective mobility increases. Note that since the effective mobility is dependent on the height of the potential barrier, both low and highly doped polysilicon (both of which have very low potential barriers) have roughly equal carrier mobility values; this agrees with experimental data [3.16].

An additional method of conduction through polycrystalline silicon is the tunneling of carriers through the top of the depletion region potential barrier. Near the tops, these barriers are sometimes narrow enough so that tunneling is likely for those carriers which possess an energy slightly less than that of the height of the potential barrier. This combination of thermionic emission and tunneling is called *thermionic field*

emission, and contributes significantly to total current only when dopant concentration is high enough (10^{17} - 10^{18} cm⁻³), or the temperature is low enough that only few carriers have more energy than that of the potential barrier.

Grain boundaries, in the basic carrier-trapping model, are assumed to be very narrow regions whose only effect is to trap carriers and create potential barriers. The disordered nature of the grain boundary can however, in of itself act as an obstacle to current flow. Restriction of current flow is due to the creation of a high, narrow potential barrier at the boundary, as Fig. 3-14 illustrates. This high, narrow barrier is a result of the physical properties of the grain boundary. The width of the barrier is equal to the physical width of the grain boundary region (a few atomic spacings), while the bandgap of the region is wider than that of single-crystal silicon and resembles that of amorphous silicon. In order to move from one grain to another, carriers must either have sufficient thermal energy to surmount the grain boundary barrier or tunnel through it. It is thought that this narrow, grain-boundary potential barrier exists in addition to the smaller, depletion region barriers which exist on either side of the grain boundary. When the composite barrier is considered, a significant portion of carriers must rely on thermionic

Figure 3-14: To model the impedance intrinsic to the grain boundary, a high, narrow barrier at the boundary is considered. The high barrier is on the order of 1 nm thick and, together with the boundary formed by depletion regions straddling the grain boundary, forms a composite barrier.

emission to overcome the barrier. It is the high, narrow grain boundary barrier which is responsible for lowering the mobility of undoped and highly doped polysilicon compared with that of monocrystalline silicon. In both undoped and highly doped polysilicon, depletion region potential barriers are either small or nonexistent and the only barrier to conduction is the grain boundary barrier.

Just as the potential barriers discussed previously produce a repulsive force on majority carriers, minority carriers see an attractive potential. These potential barriers therefore attract minority carriers, which are often caught and brought close to the defect sites. Because these defect sites act as efficient recombination centres, the minority carrier lifetime in polycrystalline silicon is significantly less than that of single-crystal silicon.

There are a couple of methods for estimating the effective minority carrier lifetime τ_{eff} . One expresses the lifetime in terms the potential barrier height V_B of the depletion region [3.24]; for the sake of brevity the derivation is left out, but it gives τ_{eff} as:

$$\tau_{eff} = \frac{2d \exp\left(-\frac{qV_B}{kT}\right)}{3\sigma v D_T (E_{fn} - E_{fp})} \quad (3-37)$$

where D_T is the grain boundary trap density (in units of $\text{cm}^{-2} \cdot \text{eV}^{-1}$), σ is the capture cross section of the defect site, d is the grain size, v is the thermal velocity of the carriers, and E_{fn} and E_{fp} are the quasi-Fermi levels for electrons and holes, respectively. Another method considers the recombination sites as being distributed uniformly throughout the polysilicon [1.14]. By this approximation the lifetime τ_{eff} is estimated to be:

$$\tau_{eff} = \frac{1}{\sigma v N_{sr}} \quad (3-38)$$

where N_{sr} is the effective density (per unit volume) of recombination centres, which defect sites act as. In the case of cubic grains, the volume density of recombination centres is given by $N_{sr} = 6N_{it}/d$ where N_{it} is the number of recombination centres per unit area of grain boundary. Both of these methods of estimating carrier lifetime show that it is inversely related to the density of recombination centres or trapping states.

Using the latter of the above two approaches, carrier lifetime can be determined from diffusion length measurements, from reverse bias characteristics of $p-n$ junction diodes, or from measuring the temporal response of a photocurrent. Since the ratio of grain boundary area to grain volume depends on the grain size, carrier lifetime also depends on grain size. Measurements taken [1.14] on carrier lifetime versus grain size showed the predicted linear dependence of carrier lifetime on grain size and led to the formulation of the empirical relation $\tau_{eff} = (5 \times 10^{-4} \text{ s/m})d$, which predicts the carrier lifetime in polysilicon with an average grain size of 50 nm to be 25 ps. This is an extremely short length of time. It is only long enough for carriers in polysilicon to travel a maximum distance of 2.5 μm , and much less than that if the fields are not strong enough to cause velocity saturation in the carriers.

To have a good performing semiconductor photodetector, the carrier lifetime must be long enough for a carrier to travel from where it is created to the nearest electrode to which it is attracted. For this reason it is not desirable to make a photodetector using a semiconductor that has so many trapping states that the carrier lifetime within that semiconductor is not long enough to allow a significant fraction of the photogenerated

carriers to travel through the semiconductor to their attracting electrodes. Therefore, a method of decreasing the efficiency of trapping states in polysilicon is required if it is to be a useful photodiode material.

3.2.4 Passivation of Trap States

As trap states in polycrystalline silicon grain boundaries adversely affect the material's electrical properties, it is often useful to remove or alter the trap states. This can be done in one of two ways: either remove the grain boundaries and therefore the traps as well, or reduce the efficiency of the trapping process. Because the former solution has the effect of significantly reducing the optical absorption of the polysilicon, it is an undesirable solution when the material is to be used as an optical detector. The latter solution therefore, is the one that has been more thoroughly considered.

Trap efficiency is reduced by *passivating* the material defects so that they do not provide active levels between the conduction and valence bands that act as trap states. Passivation of polysilicon is performed by terminating the dangling bonds that exist at grain boundaries and are due to the lattice discontinuities which form where differently oriented crystallites join. Hydrogen atoms are used to terminate the dangling bonds in solid-state silicon [3.25-3.34] and reduce the number of active trapping states. The benefit of a reduced concentration of traps is that the potential barrier height of the depletion region is reduced [3.26], thereby increasing carrier mobility and lifetime.

Several techniques exist which introduce atomic hydrogen into polysilicon. The first is to expose the polysilicon to hydrogen plasma at moderately high temperatures (300 - 400 °C); such a plasma is created by introducing hydrogen gas to a plasma reactor

[3.27]. Hydrogen plasma is essentially a gas of hydrogen atoms; these atoms are then able to diffuse into a polysilicon sample and bond to the grain boundary defect sites, thereby passivating them. Another way to add atomic hydrogen is by diffusing it from a plasma-deposited nitride layer that contains hydrogen; this type of nitride layer is often used to passivate integrated circuits [3.28]. Annealing the polysilicon at low temperatures is necessary to allow the atomic hydrogen to diffuse in. Finally, hydrogen can also be added by first implanting it into the polysilicon and then applying a mild heat treatment to cause the hydrogen atoms to diffuse to the grain boundaries, where they can react with defect sites and eliminate their trapping ability. Also, lithium can be used instead of hydrogen to passivate grain boundary sites [3.29].

While the passivation of polysilicon by atomic hydrogen eliminates many of the trapping states, it by no means removes all of them. And those defect sites that are passivated can lose the hydrogen atom that passivates them by annealing them at temperatures in the range of 400 - 500 °C [3.30, 3.31]. To maintain the passivation then, it must be performed after all high-temperature processes in the device fabrication are complete. The passivation process is however repeatable, so that if it is somehow lost the process can be successfully performed again.

CHAPTER 4

DETECTOR STUDY AND EVALUATION

4.1 Introduction

For advanced, high-speed data communications in high-performance digital networks, it is desirable to use optical interconnections. In long-distance communications, where optical signal loss and dispersion is significant, the wavelength of choice is either 1.55 μm or 1.3 μm where optical fiber attenuation is minimal. Systems that operate at these wavelengths are designed to achieve the highest possible performance and are therefore expensive. In short-distance data links such as chip-to-chip, board-to-board, and computer-to-computer connections however, where loss and dispersion are less significant, cost determines the wavelength of use. For short-distance communications, 850 nm is the best choice because lasers and detectors for this wavelength are the most cost-effective and are readily available. Likewise, it is desirable to use a low-cost photoreceiver in such data links. Cost reduction is critical to ensuring increasing use of optical data links in applications having lower required bandwidth, transmission distances, and profit margins. The low-cost, high-reliability, and proven manufacturability of silicon integrated circuitry make it ideal for use in such applications.

From a cost perspective, it is advantageous to have a photoreceiver that can be integrated with silicon VLSI circuitry. There are many potential cost advantages associated with integrating optoelectronic devices with silicon VLSI technology (optoelectronic integration). The advantages include the ability to be produced in high volume quantities with proven reliability, and the ability to combine multiple functions onto a single chip. Another cost advantage of optoelectronic integration lies in the area of optical alignment, one of the high-cost aspects of high-performance optical links where small-area photodetectors are the norm. One of the largest sources of receiver noise is the shunt capacitance at the receiver input [4.1], and an integrated photodetector can eliminate much of the parasitic capacitance at the receiver input. With most of the parasitic capacitance gone, an integrated photodetector can be made larger while contributing the same shunt capacitance as a discrete photodetector. The larger integrated photodetector will save system cost by lowering the challenge of optical alignment to the photodetector. Costs incurred from the testing, processing, packaging, and mounting of a separate chip can all be saved by integrating the receiver of an optical link with the VLSI circuitry it is intended to work with. Since the vast majority of VLSI circuitry is processed on silicon wafers, a low-cost photoreceiver that can be integrated with silicon VLSI circuitry is desirable. To this end, a silicon VLSI-compatible photodetector is required; such a photodetector has been developed as part of this thesis work. Device and system cost is very difficult to predict when the manufacturing facility, production volume, performance criteria and tolerances, testing methodology, and so on are all unknowns, however it is the estimate of the author that a receiver cost reduction of a factor of 1000 is possible with an integrated silicon photodiode.

The photodetector that was developed is a Metal-Semiconductor-Metal type (see chapter 3). The advantages of MSM photodiodes are numerous, they offer: planarity; simplicity of fabrication; and low per unit area capacitance. The fact that they are planar makes monolithic integration with VLSI circuitry easier than with a layered device such as a *p-i-n* diode. It is possible to form the interdigitated electrodes of a MSM photodiode with the same mask and metallization used to make contacts for transistors, resistors, and other devices found in VLSI circuitry. The benefit of monolithic integration, aside from reduced packaging costs, is the elimination of the bond wire between photodetector and receiver. The interaction of a discrete photodetector with receiver circuitry suffers from the length of wire that couples the two together. This wire adds an appreciable parasitic inductance when the photodiode is pulsed at high speeds, which can distort the signal and lower the bandwidth of the system.

The fact that MSM photodiodes have a lower capacitance per unit area than most other types of photodetectors, including *p-i-n* diodes, means that they can be made larger without increasing the capacitance seen on the input of the receiver. Ironically, a large portion of the cost of optoelectronic components is due to the packaging issue of optical alignment. Large area photodetectors allow the critical optical alignments to fibres or waveguides to be relaxed. In this way, the use of MSM photodiodes can reduce the cost of optical links.

The disadvantages of MSM photodiodes are: 1) that the responsivity of MSM photodiodes is poor compared with *p-i-n* photodiodes; 2) that much finer feature sizes are required for MSM photodiodes than are required for *p-i-n* diodes; 3) that the metal-semiconductor interface can be hard to control reliably; and 4) that their operating speed

can be limited by the slow response of carriers generated in low electric field regions deep under the electrodes. Disadvantage 1 will be addressed later on in this thesis, and is also being addressed by others [2.12, 4.3]. The minimum feature size required for MSM photodiodes is much smaller than that required for *p-i-n* diodes, the desirable minimum feature size is of the order of the absorption depth of the wavelength of light that is used in the absorbing material that is used. While this dimension ranges from several microns down to several tenths of microns, these minimum feature sizes can be achieved with modern VLSI technology, and so disadvantage 2) is rapidly becoming invalid. Disadvantage 3) is a minor one, but one that nevertheless requires further study to overcome. Disadvantage 4) will be addressed later on in this thesis as well as presently being addressed by others [1.15, 2.6, 4.4].

The responsivities of MSM diodes are lower than those of *p-i-n* photodiodes. MSM responsivities are lower due to 1) opaque electrodes which block a portion of the incident light from the photosensitive material below them; and 2) finite carrier lifetimes due to recombination centres which cause carriers to recombine before reaching the electrodes. Research into the use of transparent electrodes however, such as indium tin oxide (ITO) or cadmium tin oxide (CTO), is underway [4.2]. These electrodes transmit incident light while conducting the photocurrent produced by the photodiode. The effect of shadowing by the electrodes is in this way reduced. Thus far however, their reliability is questionable. Finite carrier lifetimes can be increased by techniques such as hydrogen passivation of recombination centres [3.25-3.34], and inter-electrode spacing can be reduced so that carrier transit time is less than carrier lifetime. In either case, the fraction of carriers which reach the MSM electrodes will increase.

It is however, possible to exploit the finite carrier lifetime. The finite carrier lifetime can be used to enhance the speed of a MSM photodiode by eliminating a percentage of the free carriers after a certain time. In this way, the slow moving carriers recombine after a time and are therefore unable to produce an extended temporal response tail. And, if high speed is required at the expense of responsivity, it is possible to make an absorption region having many defect sites which act as recombination centres for free carriers. The slower carriers will then recombine before reaching the device electrodes, resulting in poor device responsivity but high-speed performance.

One of the most challenging problems associated with MSM photodiodes is the difficult to control nature of metal-semiconductor interfaces. This issue stems from pinning of the Fermi level at the metal-semiconductor interface by surface states (see chapter 3). Silicon, however, is known to have very few surface states and has well-controlled interfaces with metals [3.2, 3.5]. For other semiconductors, such as the III-V semiconductors, work is being done on passivating the semiconductor surface to satisfy the dangling bonds present there. Also, new methods of mass wafer processing are being explored, where the active region of the wafer is grown in a vacuum environment and then transferred directly to a metallization chamber without leaving the vacuum. In this way, surface contamination cannot occur.

Another problem that plagues MSM photodiodes is the long temporal response tail caused by the photogeneration of carriers in low field regions. These carriers drift slowly towards one of the electrodes until they reach high field regions, at which point their speed approaches their saturation velocity. To eliminate the effect of slow carrier diffusion, the absorption region can be made thin enough so that all areas of the

absorption region have a sufficiently high electric field. The MSM photodiode presented here has an absorption region that is only 2 μm thick. For a good compromise between device speed and responsivity, the absorption region should only be thick enough to absorb 60% to 85% of the light incident upon it. Trying to absorb too much of the incident light with a thicker absorption region leads to extended tails in the temporal response curve of the device, which greatly reduce the speed of the detector. Carriers created beside the detector can also produce an extended temporal response tail to some extent. These carriers can be eliminated by performing a mesa etch around the MSM photodiode, so that unwanted semiconductor around the detector is removed.

The MSM photodiode that was developed uses polysilicon as the photoactive material. Polysilicon is known to have a higher absorption coefficient than monocrystalline silicon (chapter 3), for this reason the photoactive material layer can be made thinner than it would be if it were composed of monocrystalline silicon, and yet still absorb an equivalent amount of optical energy. The reason for increased absorption in polysilicon, as discussed in section 3.2, is the presence of short-range disorder at the crystallite boundaries. The structural defects present at grain boundaries act as absorption sites, but also act to impede the flow of charge carriers as discussed in section 3.2.3. In this way a tradeoff exists between gain and bandwidth when polysilicon is used as the photoactive material in a photodiode.

The tradeoff that exists between gain and bandwidth is one reason why it is beneficial to have a photoactive material that has a high optical absorption coefficient. A material with a higher absorption coefficient will absorb photons more strongly and will not allow them to penetrate as deeply into the bulk of the material, equation 2-8 shows

this relationship. By absorbing incident photons near the surface, the photogenerated carriers will be created in a region of high electric field [2.5, 4.7]. Figure 4-1 illustrates typical equipotential lines formed in a planar MSM device. Electric field lines run perpendicular to equipotential lines and their magnitude is directly proportional to the density of equipotential lines. Thus, within the semiconductor of the MSM device, electric field strength is highest near the surface where the metal contacts are, and become weaker as depth increases. When the photogenerated carriers travel towards the electrodes, a displacement current is created in the device external circuit. It is possible to notice a perturbation of the internal electric field generated by the external bias when the level of illumination is sufficiently large. The high level of illumination incident on a photodiode generates enough charge carriers within the semiconductor to screen the internal field, causing the peak amplitude of the pulse response to increase at a less than linear rate and the tail of the response to increase substantially [4-18].

Figure 4-1: Equipotential lines in a planar MSM structure, the heavy lines represent metallic contacts.

High electric fields cause photogenerated carriers to drift quickly toward electrical contacts, and thereby allow detectors in which carriers are produced in high-field regions to have a fast response. It is important therefore to minimize the fraction of carriers that are generated in low-field regions. Methods of achieving high-field absorption of photons include the use of materials (GaAs, InGaAs, InP) which possess a high absorption coefficient, and by using an optical wavelength which is sufficiently short (i.e. 400 nm) to ensure a small absorption depth in Si (100 nm at $\lambda = 400\text{nm}$) [1.27]. As well, the use of a thin layer of monocrystalline silicon at an optical wavelength of $\sim 800\text{ nm}$ where the absorption depth is $\sim 10\text{ }\mu\text{m}$ can produce high-field absorption. Recently, methods requiring advanced wafer processing capabilities have been used to create a thin layer of monocrystalline silicon on which to build MSM photodiodes [1.15, 4.4].

The approach taken in this work differs somewhat from those taken in the past to build a silicon-based MSM photodetector. Rather than using exotic optical wavelengths or wafer processing methods, a $2\text{ }\mu\text{m}$ thick layer of polycrystalline silicon (polysilicon) was used to limit the absorption depth of incident photons, thereby ensuring photogenerated carrier creation in high-field regions. As previously mentioned, the other benefit of using polysilicon as the photoactive material is the higher absorption offered. The disadvantage of using polysilicon lies in its low mobility, as transit times for carriers in polysilicon can be much greater than for those in monocrystalline silicon.

4.2 Device Fabrication

The detectors were fabricated on readily available (100)-oriented, single-side polished silicon wafers. Equipment limitations necessitated the use of a lift-off process in patterning the metallization layer. A standard process of selectively etching away unwanted metal through the use of a patterned photoresist layer was unsuccessfully attempted several times before the lift-off process was developed. Lift-off processes are generally more difficult to accomplish than a selective etch process. A description of the lift-off process follows; as well, an illustration of the process is given in Figure 4-2. The silicon wafers used initially had a 100 nm layer of SiO_2 deposited on their top surface, this layer acted to insulate the photoactive layer from the substrate. On top of the oxide layer, a 2.0 μm thick layer of undoped polysilicon was deposited. The polysilicon was grown using low pressure chemical vapour deposition (LPCVD) at a temperature of 627 °C and at atmospheric pressure from a SiH_4 source. A lower layer of photoresist (HPR504) was then spin coated onto each of the wafers at 4000 rpm for a duration of 30 seconds, following which a 100 nm thick layer of Al was deposited by evaporation. The next layer applied was the top layer of photoresist (again HPR504 spin coated at 4000 rpm for 30 seconds); it was this layer that initiated the pattern transfer from the mask that carried the original electrode pattern. Pattern transfer to this top photoresist layer was accomplished by the standard method of exposure and development, leaving behind a positive copy of the mask pattern. The wafers were then sputter etched at 80 W for 18 minutes to remove all traces of Al in regions which were left exposed by the windows formed in the overlying photoresist layer. In this way, the pattern was transferred from the top photoresist layer to the Al layer. The next step was to plasma etch the wafers in

(a) Deposit layers of oxide, polysilicon, photoresist, and Al

(b) Expose and develop top layer of photoresist

(c) Sputter etch Al through windows in top photoresist

(d) Plasma etch exposed photoresist and deposit metal layers

(e) Dissolve photoresist and lift-off unwanted metal in acetone

Figure 4-2: Basic steps in the lift-off procedure for fabrication of polysilicon MSM photodiode.

an O₂ environment. This etch was performed for a total of 60 min at a pressure of 0.500 T and at a power of 300 W. An O₂ plasma etch is necessary to remove the bottom layer of photoresist in areas left exposed by windows formed in the above Al layer, further transferring the mask pattern. A short (30 second) sputter etch to clear away any remaining photoresist was then performed on the wafers before metal deposition. The metal layer consisted of a sandwich of three metals: Ti, TiW, and Al. Ti and TiW were sputtered on in layers of 10 nm and 40 nm respectively. The layer of Al was the last deposited, it was evaporated on forming a layer having a thickness of 300 nm. Following metal deposition, the wafers were placed in an acetone ultrasonic vibration bath. The purpose of the bath was to dissolve the bottom layer of photoresist and detach or lift-off the overlying metal layers, leaving a negative copy of the mask pattern on the wafer surface formed by the deposited metal layers.

Titanium was chosen as the bottom metal (contact metal, or Schottky metal) because, as previously discussed, its work function produces electron and hole Schottky barriers of approximately equal heights when both anode and cathode are deposited using similar processes, and it also has good adherence to silicon. Since both anode and cathode are deposited simultaneously in identical conditions, the two barrier heights should be approximately equal; this should minimize device dark current. Aluminum was used as the top metal because of its ability to adhere well to gold bond wires, which are used to carry device photocurrent off-chip. Titanium tungsten alloy was used between the Ti and Al layers to keep Al from diffusing into the silicon.

The anode and cathode of the photodiode form an interdigitated structure. Each electrode consists of ten parallel fingers joined together at one end. The fingers of each

electrode have been placed in an alternating sequence consisting of anode followed by cathode. The fingers forming the anode are joined together at one side of the device, while the fingers of the cathode meet at the opposite side. A photograph of one of the devices is presented in Figure 4-3.

Following metallization, wafers were annealed in a hydrogen atmosphere. The effects of annealing polysilicon in hydrogen plasma have been well documented [1.14, 3.24]. As discussed in chapter 3, hydrogen can bind with dangling bonds present at crystallite boundaries in polysilicon, effectively eliminating those that are able to bind with a hydrogen ion. The effects of hydrogen passivation include increased carrier mobility and lifetime, as discussed in chapter 3. Two methods of hydrogen passivation

Figure 4-3: Photograph of polysilicon MSM photodiode showing the interdigitated structure of the anode and cathode.

were performed: one was to simply anneal the wafers in a molecular hydrogen atmosphere, while in the other method, hydrogen plasma was used. Annealing in a molecular hydrogen atmosphere was performed at 400 °C and at atmospheric pressure. Test wafers had 500 nm of polysilicon deposited on them, and after metallization were annealed for durations of 45 minutes, 90 minutes, and 180 minutes respectively. Table 4-1 shows the measured photocurrent and dark current under identical temperature and illumination conditions. Focussing the light from an incandescent bulb onto the detectors provided illumination, and a bias of 5 V was supplied to each of the detectors. Based on the information in Table 4-1, detectors having a full 2.0 μm of polysilicon were fabricated and subsequently annealed in a molecular hydrogen atmosphere for 100 minutes in an effort to maximize responsivity. A 90 minute anneal, in accordance with the times listed in Table 4-1, was attempted but the wafers were accidentally left in the oven for an extra 10 min. As the standard deviations for these data in Table 4-1 are relatively high, more wafers would need to be produced and devices tested to establish reproducibility of photocurrent and dark current values.

Another method of introducing hydrogen and eliminating dangling bonds present at the crystallite boundaries of polysilicon was tested. A second group of wafers were annealed in hydrogen plasma created within a parallel-plate plasma chamber [3.27, 3.28, 3.30 – 3.33]. These wafers were annealed for 45 minutes at a temperature of 320 °C, a pressure of 1.0 Torr, and a power of 200 W over the 9 inch diameter parallel plates of the chamber. This passivation in hydrogen plasma served only to increase the dark current of the devices and was not used to make the final detectors. It is suspected that the wafers were contaminated when annealed in the plasma chamber. However, due to the large

number of successful passivation treatments on polycrystalline silicon reported in the literature [3.27, 3.28, 3.30 – 3.33], it must be noted that another attempt at passivating with a hydrogen plasma should be undertaken. Unfortunately, limited access to a plasma chamber prevented a repeat of this procedure.

Table 4-1: Measured photocurrent and dark current from test detectors annealed for various times in a molecular hydrogen atmosphere.

Duration of hydrogen anneal (minutes)	Photocurrent (nA)		Dark Current (nA)	
	mean	Standard Deviation	mean	standard deviation
0	480	51	424	49
45	570	62	210	22
90	720	84	250	27
180	550	60	180	18

4.3 Device Model and Simulation

In order to understand better the operation of the device, a model of the detector was produced and then used to simulate the device. The goal was to keep the model as simple as possible while maintaining a close resemblance to the physical properties of the actual device. One such model chosen was a 2-dimensional one in which a layer that had properties characteristic of polysilicon represented the absorbing region of the device. A 1-dimensional model was also used for comparison. An absorption coefficient equal to the measured value found for our polysilicon samples and reported in chapter 3 was used

in the model, and the relative permittivity used was that of Si ($\epsilon_r = 11.7$), while carrier mobility and lifetime were varied.

By taking advantage of symmetry, only 2 metal contacts were used to represent cathode and anode, respectively. The contacts used in the 2-dimensional model were perfect conductors of $1.25\ \mu\text{m}$ width, with $2.5\ \mu\text{m}$ separating adjacent contacts. Mirror symmetry was assumed to exist on either side of an axis drawn vertically through the center of each contact. For this reason each contact was modeled with a width equal to half of the full width of each contact finger used in the actual device. Polysilicon thickness was varied from 1 to $3\ \mu\text{m}$ in the model. An illustration of the model device structure is given in Figure 4-4. The metal-semiconductor interface was made to have a Schottky barrier height of 0.50 eV to simulate the estimated Schottky barrier of the actual detector. Beneath the polysilicon (in the model), an insulator of infinite thickness with a dielectric constant equal to that of Si was assumed. This insulator represented the thin layer of insulating oxide over the bulk silicon of the wafer that the device rests upon. While the insulator used in the model does not accurately represent the thickness of the insulator used in the actual device it does accurately model the combination of a thin (100 nm) insulator on top of bulk silicon, and was used in order to simplify the model. The insulator used in the model has a relative permittivity equal to that of silicon and so no discontinuity in dielectric constant exists across the polysilicon/bulk insulator interface. This closely models what we expect in the actual device where only a negligible discontinuity in dielectric constant occurs at either side of the 100 nm thin oxide layer. In the model, diffraction of the incident optical beam by the electrode pattern was not

accounted for since diffraction effects are negligible in the near field, and we are only interested in the near field after the electrode pattern.

Figure 4-4: Illustration of 2-D model used to simulate detector operation. Symmetry allowed use of only 2 half-width contact fingers. Some electric field lines are shown.

The 2-dimensional device model was simulated using Medici [4.5] software. Several assumptions were made in the simulation. The model was two-dimensional, and so the electric fields in it were also two-dimensional. The electric field was therefore assumed to be everywhere perpendicular to the long axis of the metal contact fingers. Thermionic emission over a Schottky barrier was assumed to be the only method of conduction through the device in the absence of illumination. Scattering of carriers by phonons, electrons, and impurities are all assumed to exist and are accounted for by defining electron and hole mobility values that fit with experimental data. Parasitic effects such as inductance and edge effects are neglected. The optical pulse which illuminates the device in the simulation is a piecewise linear function of time that resembles the shape of the 90 ps FWHM wide laser pulse used to test the actual device.

The illuminating optical pulse and the simulated response of a device having a 2 μm thick polysilicon layer, and a 12 V applied bias are both given in Figure 4-5.

The 1-dimensional device model consisted of a layer of polysilicon that was sandwiched between two Schottky contacts. A somewhat thicker polysilicon layer is needed to emulate the average path length for carriers generated in the 2-dimensional model. Assuming that 60% of the light that reaches the polysilicon layer is absorbed before passing through it and into the underlying substrate, then the depth at which 30% of this light is absorbed represents the average depth of light penetration. Given that the optical absorption coefficient for $\lambda=850\text{ nm}$ light is $4.0 \times 10^3\text{ cm}^{-1}$ for polysilicon as presented in chapter 3, the average penetration can be calculated to be $0.89\text{ }\mu\text{m}$. The paths that carriers take through the polysilicon in the 2-dimensional simulation

Figure 4-5: Two-dimensional Medici simulation of poly MSM photodiode showing incident optical pulse (top) and simulated device photoresponse (bottom).

can be approximated as the arcs of a family of ellipses, the centers of which lie on the top surface of the polysilicon, equidistant between the two contacts, and the low points of which lie at various distances directly beneath the mid-point between the two contacts. Arcs such as these are illustrated in Figure 4-4. The equation of an ellipse is $x^2/a^2 + y^2/b^2 = 1$ in Cartesian space, where x and y measure distance along mutually perpendicular axis as shown in Figure 4-4. The terms a and b are the x and y -intercepts, respectively. The y -intercept b is the depth of light penetration, and the x -intercept is approximated as $a = 0.5(L + W - Le^{-b})$, where W and L are the contact finger width and inter-finger distance respectively. The circumference C is approximated for simplicity as [4.6] $C = 2\pi[(a^2+b^2)/2]^{1/2}$. As the elliptical arcs shown in Figure 4-4 are each one-half of a full ellipse, their path length is one-half of the circumference of a complete ellipse. When $W = L = 2.50 \mu\text{m}$ as is the case for the 2-dimensional simulation, the estimated average path length is $4.14 \mu\text{m}$.

In this simplistic model, it is assumed that device response is limited by both carrier transit time and carrier lifetime; RC time constants are neglected. It is also assumed that photogenerated carriers are uniformly created between the two contacts. Carrier transit effects produce a linear photocurrent pulse response as carriers exit the semiconductor through the contacts at a constant rate. Carrier lifetime effects alone will cause the photocurrent pulse response to have an exponential decay. When both carrier transit and lifetime are considered, the photocurrent pulse response can be obtained by multiplying the linear response of the transit effect by the exponential response of the carrier lifetime effect. When carrier transit time is much greater than carrier lifetime,

then lifetime will dominate the pulse response, and when transit time is much less than carrier lifetime, transit time will dominate.

Carrier mobility values were varied to determine the effect on device response. The optical pulse that illuminates the detector in this simulation is simply a delta function in time. The applied bias voltage in this simulation was again 12 V, just as it was in the 2-dimensional simulation. Polysilicon layer thickness in this simulation was $4.14\text{ }\mu\text{m}$, as determined by estimating carrier path length. Results of the 1-dimensional simulation are illustrated in Figure 4-6. The two distinct constant slopes of the device photoresponse are a result of the separate electron response and hole response, both of which are linear functions of time. Because electrons have a higher mobility than holes and therefore a shorter transit time, the electron response curve has a higher peak current and a shorter response time than the hole response curve. This effect causes a discontinuity in the slope of the total response curve to occur where the value of the electron response

Figure 4-6: One-dimensional simulation of the device using only carrier transit time to determine device response. Electron and hole responses are shown in dashed lines. The stimulus in this simulation is an optical pulse that is a δ function in time, which produces a uniform distribution of charge carriers throughout the polysilicon region. Electron mobility is $26\text{ cm}^2/\text{V}\cdot\text{s}$, and hole mobility is $13\text{ cm}^2/\text{V}\cdot\text{s}$.

reaches zero. The 1-dimensional simulation clearly shows that photocurrent peak response is due mainly to electrons, while the relatively low speed of holes produces the tail of the device response curve.

Comparing the response curves from the 1-dimensional and 2-dimensional simulations, it may be seen that the full-width-half-maximum (FWHM) of the 2-dimensional simulation is 260 ps, while the FWHM of the 1-dimensional simulation is 250 ps. The 2-dimensional Medici simulation is complex and takes into account many factors, while the 1-dimensional simulation takes only transit time across a single straight-line path into account and yet is capable of predicting a response curve similar to that of the 2-dimensional simulation. Figure 4-7 shows a Fourier transform of the response curve from the 2-dimensional simulation. It was performed in order to produce a simulated device frequency response, and it shows that the -3 dB frequency point is 1.2 GHz. For comparison, the Fourier transform of a response curve similar to that shown in Figure 4-6 can be found in Figure 3-10.

Figure 4-7: Frequency response of simulated 2-dimensional MSM device. The curve is the fast Fourier transform of the data given in Figure 4-5.

4.4 Experimental Results

Polycrystalline MSM photodiodes were characterized at DC with a continuous-wave laser light source as well as at high frequencies using a pulsed laser to stimulate the detector, along with a high-speed probe station and oscilloscope to collect and display the response. To characterize the photodiodes at both DC and high frequency, light from a laser diode was incident on the detectors via a standard multimode optical fibre having a core diameter of $62.5\text{ }\mu\text{m}$. The incident light was normal to the plane of the detector, and was positioned by placing the optical fibre in a micro-positioner that was operated manually while the detector and fibre end were viewed through a microscope. The arrangement is illustrated in Figure 4-8. In this way, the fibre could be placed such that the core of the fibre was centered several microns over the photodiode, so that the amount of light from the fibre that was incident on the detector could be maximized when responsivity measurements were taken.

Figure 4-8: Illustration showing bare and cleaved fibre end positioned directly above photodiode. This setup was used in both DC and high frequency characterization of the polysilicon MSM photodiodes.

To characterize DC operation of the detectors, a Mitsubishi model FU-21LD-SMF laser diode operating at $\lambda = 860$ nm was used to provide light to the detector via the optical fibre as mentioned above. Optical power emitted through the end of the optical fibre from the laser diode was measured with a Photodyne 22XL optical multimeter and was maintained at a constant level by monitoring the current output from the built-in *p-i-n* photodiode contained within the assembly of the FU-21LD-SMF laser diode. The built-in photodiode measures light output from the back face of the laser diode, which ideally is equal in power to the light that is emitted through the front face of the laser. It is the light emitted through the front face that is used in the DC characterization discussed here. Due to the close proximity of the fiber to the detector under test, it was assumed that all of the optical power emitted through the end of the fibre was incident on the detector. DC current output from the polysilicon MSM photodiodes was monitored with a Fluke 85 multimeter. Current versus voltage data is plotted in Figure 4-9. The

Figure 4-9: Current-voltage characteristics of a polysilicon MSM photodiode at different light intensities. The lowest-most curve is the current-voltage characteristic for no incident light, and is called the dark current characteristic.

Figure 4-10: High-speed test arrangement for polysilicon MSM photodiode showing OTDR laser source, bias-T, and oscilloscope.

measured responsivity at 12 V applied bias was 0.13 A/W, which corresponds to an external quantum efficiency of 19%. Accounting for light blocked by contact fingers and reflected by the air/polysilicon interface gives an internal quantum efficiency of 54%. Above applied bias values of 15 V, the onset of avalanche breakdown was observed. The slope of the I - V curves in Figure 4-9 is not typical of ideal MSM photodiodes and indicate the presence of low frequency gain which studies have indicated is produced by traps on the semiconductor surface [4.8].

The temporal response characteristic of the polysilicon MSM photodiodes was determined by using a pulsed optical-time-domain-reflectometry (OTDR) laser diode (Opto-Electronics PPL30K) that produced pulses with a nominal wavelength of 850 nm. The optical pulses had a FWHM of 90 ps and a pulse repetition frequency of 30 kHz. These optical pulses were incident on the photodiodes via an optical fibre as discussed above. The signal produced by the detector was observed using an RF probe and a 4.2 GHz bandwidth bias-T arrangement as illustrated in Figure 4-10. The bias-T was used to

provide a bias voltage for, and collect the detected signal from, one of the photodiode's electrodes while the other was grounded. Figure 4-11 (a) shows the response of the polysilicon MSM photodiode in the time domain for applied bias values of 4, 8, and 12V.

(a)

(b)

Figure 4-11: Response of a polysilicon MSM photodiode with $2.5\ \mu\text{m}$ finger width and spacing in both (a) the time domain, showing response curves for 4, 8 and 12 V applied bias normalized to the maximum height of the 12 V response curve, and (b) the frequency domain which was produced by computing the fast Fourier transform of the 12 V time domain response from (a).

Table 4-2: Average measured response times of 16 polysilicon MSM photodiodes at various applied bias voltages.

Applied Bias (V)	Rise Time (ps)	Fall Time (ps)	FWHM (ps)
4	77	436	231
8	84	457	239
12	91	476	255

The absence of an elongated tail in the photocurrent response indicates the absence of a high-frequency gain. This type of gain is visible in time domain pulse response curves, yet is absent from DC I - V curves. High-frequency gain is responsible for decreasing the bandwidth of photodiodes which exhibit this characteristic [4.8]. The fast Fourier transform of the time domain response data for the case of 12 V applied bias was calculated and is presented in Figure 4-11 (b). This data shows the amplitude of the frequency components present in the impulse response, and from it one can see that the impulse response has a -3 dB bandwidth of 750 MHz. The average 10%-90% rise times, fall times and FWHM values from 16 devices from a common wafer are recorded in Table 4-2. Since the fall times for the response curves produced from different bias voltages are approximately equal, it is reasonable to conclude that a common factor that is independent of bias and electric field strength limits the response of the devices. That common factor is the charge carrier lifetime of the polysilicon used in these devices.

Two-dimensional Medici simulations were performed with various values for charge carrier lifetime. The simulation result that best matched the measured response curves had carrier lifetimes of 220 ps. Further simulations were then performed, which

incorporated the carrier lifetime value of 220ps, in an effort to determine the carrier mobility values. By varying the mobility values for electrons and holes, but by maintaining that electron mobility must be twice that of hole mobility, carrier mobility values were determined by comparison of simulated response curves with the measured curves. Although carrier saturation velocity will likely vary depending on the structural parameters of the polysilicon used, carrier saturation velocity was assumed to be 10^7 cm/s, as it is for monocrystalline silicon [2.2], and close to the value of 6×10^6 cm/s that one group of researchers found for p-type polysilicon [4.9].

The effect of changing mobility values in simulation is a change in the peak of the response pulse relative to those produced at other bias voltages. Mobility values that were too high caused the peaks of the response curves produced at 4, 8, and 12 V applied bias to be too similar in value because in each case the carriers approached saturation velocity. Mobility values that were too low caused the peak of the 4 V bias response curve to be too low in relation to the peak of the 12 V bias response curve. The mobility values used in simulation that best fit the measured data were $\mu_n = 26 \text{ cm}^2/\text{V}\cdot\text{s}$ for electrons, and $\mu_p = 13 \text{ cm}^2/\text{V}\cdot\text{s}$ for holes. These values compare well with those found in the literature for polysilicon [3.15, 3.16, 3.20, 3.38]. Because of the good match between experiment and simulation using these values of carrier lifetime and mobility values as illustrated in Figure 4-12, we concluded that the carrier lifetime in our polysilicon samples was 220 ps. Similarly, we concluded that the carrier mobility values of our samples were $\mu_n = 26 \pm 2.6 \text{ cm}^2/\text{V}\cdot\text{s}$ for electrons and $\mu_p = 13 \pm 1.3 \text{ cm}^2/\text{V}\cdot\text{s}$ for holes (the error on the measured rise and fall times of the devices was $\pm 10\%$).

A widening of each of the measured response curves can be observed in Figures 4-11(a) and 4-12. The response curves appear to widen below the half-current point of each curve to produce a hump in the pulse response approximately 500 ps after the main peak. The hump has the appearance of a reflection, from a mismatched impedance, of the main peak. The time delay between the main peak of the pulse response and the hump suggest that the reflection is between the RF probe used to contact the device pads and the bias-T used in the test arrangement. The measured pulse delay in the SMA environment is 6.0 ns/m and the distance between the RF probe tips and the bias-T was approximately 4 cm. The time delay for a single reflection in this environment is then 480 ps, which agrees well with the approximately 500 ps noticed in the pulse response between the peak and hump.

Figure 4-12: Measured and simulated response curves for a 2.5 μm finger width and spacing polysilicon MSM photodiode with a 2.0 μm thick polysilicon layer on a silicon substrate. The simulated response curves were made using a carrier lifetime of 220 ps, and electron and hole mobility values of 26 $\text{cm}^2/\text{V}\cdot\text{s}$ and 13 $\text{cm}^2/\text{V}\cdot\text{s}$ were used, respectively. This figure was made by superimposing the three sets of curves and spacing them temporally to aid in viewing.

Capacitance measurements of the device were made with an HP 4280 1 MHz capacitance meter, measurements were made at 0 V bias and at 10 V bias. The capacitance meter has a resolution of 1 fF, though errors due to added cable and probes had to be accounted for. These effects totaled 72 ± 10 fF and had to be subtracted from the instrument readout in order to obtain proper measurements of the devices. For every device measured the pads and signal traces were cut away so that only device capacitance could be measured, and all measurements were done in the dark to eliminate any photoconductive effects. Capacitance values between device contacts at 0 V bias ranged from 154 ± 10 fF to 178 ± 10 fF for a total of 16 devices measured with an average value of 168 ± 10 fF and a standard deviation of 6.5 fF. For an applied bias of 10 V the measurements on 16 devices ranged from 142 ± 10 fF to 164 ± 10 fF with an average value of 153 ± 10 fF and a standard deviation of 6.6 fF. In both cases, the stray capacitance that was measured when the probe tip was not in contact with the device pad was 70 ± 10 fF. Subtracting this amount from the meter readings gives an average capacitance of 98 ± 10 fF at 0 V bias and 83 ± 10 fF at 10 V bias. The reduced capacitance at 10 V can be explained by the reduced depletion region thickness caused by the applied electric field. The device capacitance can be calculated from Equation (3-5) with $N = 19$, $l = 95$ μm , and $W = L = 2.5$ μm ; it gives $C = 101$ fF.

Capacitance measurements of device to substrate were also made. At 10 V bias, the average value of 16 measurements was 2.3 pF with a standard deviation of 90 fF, while at 0 V bias the average value of 16 measurements was 2.2 pF with a standard deviation of 85 fF. By severing the trace where it connects to the device, capacitance measurements of pad and trace to substrate could be made, the average capacitance of 16

measurements was 1.759 pF with a standard deviation of 75 fF, whereas before severing the traces the average was 2.340 pF with a standard deviation of 90 fF. The difference of 581 fF is the device capacitance to substrate. By subtracting the stray capacitance of 894 fF (measured with the probe lifted off the surface of the pad) from each of the measured values of pad plus trace plus stray capacitance, a result of 865 fF is obtained for the value of pad and trace capacitance to substrate. The calculated capacitance of an $80\text{ }\mu\text{m} \times 80\text{ }\mu\text{m}$ pad over a $2\text{ }\mu\text{m}$ thick dielectric layer with a dielectric constant of 11.9 is 313 fF by itself, while the sum total of pad and trace area gives a calculated capacitance of 500 fF. By considering fringe effects a value similar to that measured for the pad and trace can be expected.

The devices showed no noticeable loss of photocurrent or increase of dark current in intermittent testing over a period of 3000 hours. In continuous testing with a device under illumination for 123 hours, no decrease in photocurrent was measured and no increase in dark current, other than that expected from the rise in temperature due to constant illumination, was noticed.

4.5 Additional Simulations

Further simulations were performed to determine the response of similar photodiodes having polysilicon layers of different thickness. Simulations were run of photodiodes with polysilicon layer thickness of $1\text{ }\mu\text{m}$ and $3\text{ }\mu\text{m}$ along with the $2\text{ }\mu\text{m}$ thickness already simulated. These simulations were performed to study the effect of polysilicon layer thickness on device performance. Temporal responses to an optical

pulse produced by these simulations are shown in Figure 4-13; in each case the pulse produced by a 2 μm thick polysilicon layer is shown for comparison. The simulation results shown in Figure 4-13 (c) were produced by simulating a polysilicon MSM photodiode with a 3 μm thick polysilicon layer in which the carrier lifetime was increased by a factor of 10 to 2.2 ns, instead of the 220 ps used in the other simulations. The FWHM for the 1 μm thick polysilicon layer simulation was 210 ps compared with 260 ps for the simulation in which a 2 μm thick polysilicon layer was used, however the shorter FWHM was at the expense of a significantly lower peak current response as Figure 4-13 (a) illustrates. By increasing the polysilicon layer thickness to 3 μm , the FWHM and peak current response were virtually identical to that of the simulation of the 2 μm thick device. This is because the carrier lifetime of 220 ps prevents nearly all of the carriers generated at levels deeper than 2 μm from reaching the device contacts before recombining. When carrier lifetime was increased to 2.2 ns however, the FWHM of the response from the photodiode with the 3 μm thick polysilicon layer increased significantly to 350 ps without significantly increasing the peak current of the response. The increased FWHM is due to a much weaker electric field experienced by carriers generated in the deepest levels of the 3 μm thick polysilicon layer than those generated in the upper layers of the polysilicon layer [2.5, 4.7]. These carriers that are generated in low-field regions have a relatively low drift velocity and reach the device contacts much later than carriers generated in high-field regions. As a result, the carriers generated in low-field regions contribute to the temporal response later in time than other carriers, and hence produce an extended tail in the temporal response.

Another mechanism for the generation of extended tails in the temporal response of a photodiode is carrier tunnel injection from the cathode induced by traps in surface states or deep traps very close to the contacts [4.8]. When investigating pulse response tails, researchers have found that the tails were insensitive to bulk defect concentration whereas traps near the semiconductor/metal interface have an appreciable effect on these tails [4.8]. The same researchers noticed that reducing the bias voltage in their MSM photodiodes reduces the tail amplitude as well. By decreasing the applied bias and

Figure 4-13: Two-dimensional Medici simulations of polysilicon MSM photodiode with polysilicon layer thickness of: a) 1 μm , b) 3 μm and c) 3 μm . In c) the carrier lifetime was increased from 220 ps to 2 ns. In all three cases, the fainter curve corresponds to the pulse produced by a simulated polysilicon MSM having a 2 μm thick polysilicon layer.

therefore the electric field inside the detector, the Schottky barriers at the contacts are increased both in height and width. This supports the tunnel injection theory. Carriers are able to tunnel through the Schottky barrier of a MSM photodiode if trap states exist a short distance away from the metal surface. According to quantum theory, the fraction of electrons that do tunnel through the barrier is proportional to the distance separating the trap from the metal surface. Experiments performed on MSM photodiodes at various temperatures [4.8] showed that the tails were insensitive to temperature. This suggests that trapping and detrapping effects, which vary with temperature, do not affect the photocurrent tail. Traps in the bulk material therefore have no effect on the tail of the response. Extended tails in the temporal response of other photodiodes have previously been reported [1.27, 3.5, 4.8, 4.10]. They can be particularly bothersome in monocrystalline Si where the carrier lifetime can be as high as $10\ \mu\text{s}$ [1.14], making the fabrication of high-speed Si-based detectors difficult. Typically, Si based MSM detectors are prone to extended tails due to low-field carrier generation caused by carrier generation deep under the Si surface. The large absorption depth is due to the weak absorption coefficient of Si. GaAs and InGaAs based detectors, on the other hand, are prone to electron injection from the cathode. It has been observed that GaAs and InGaAs are prone to having significantly more surface states than Si [3.5, 4.8, and 4.10].

To investigate the performance of a polysilicon MSM photodiode produced with modern sub-micron photolithography, a Medici simulation of a photodiode having $1\ \mu\text{m}$ contact finger widths and spacings was performed. Such contact finger dimensions are unattainable with the equipment available for this project so they could only be explored by way of simulation. For this simulation, the optical pulse that stimulates the detector

had to be made much shorter in duration than those used in the previous simulations. A 9 ps FWHM scaled down version of the optical pulse used in previous simulations was employed for this simulation. The need for a shorter duration pulse stems from the shorter distance between contacts used in this simulation. Smaller spacings between contacts mean higher electric fields for an equivalent applied bias, and therefore higher carrier velocities. The effect of lower inter-contact spacing on carrier transit time is therefore twofold: it is lower because transit distance is lower, and also because carrier velocity is higher. At low values of electric field, where drift velocity is linear with the electric field, carrier transit time is proportional to the square of inter-contact spacing. As electric field increases however, and drift velocity saturation begins, carrier velocity is no longer directly proportional to electric field and carrier transit time in this case is proportional to inter-contact spacing raised to the power n , where $1 < n < 2$.

To prevent the photogeneration of carriers in low-field regions, the polysilicon layer in this simulation was made to be 1 μm thick, rather than 2 μm as used in previous simulations. Simulation results are given in Figure 4-14; they indicate that a polysilicon MSM photodiode with 1 μm contact finger widths and spacings can produce current pulses having a FWHM of 56 ps and a bandwidth of 4.5 GHz. Such a bandwidth is substantially higher than the 750 MHz reported for the 2.5 μm width and spacing polysilicon MSM presented earlier in this chapter, and merits consideration as a future research project.

Figure 4-14: Simulation results of a polysilicon MSM photodiode with 1 μm wide contact fingers and inter-finger spacings, and a 1 μm thick polysilicon layer. (a) Shows the temporal response (heavy line) to a simulated incident optical beam with FWHM of 9 ps. The previous simulation results for photodiodes having 2.5 μm finger widths and spacings and polysilicon layer thicknesses of 2 μm (solid line) and 1 μm (dashed line) are also given for comparison. The simulated bandwidth (b) is also given, it shows that this detector has a simulated -3 dB bandwidth of 4.5 GHz.

CHAPTER 5

OPTICAL RECEIVER

5.1 Introduction

Receivers for local optical interconnections differ in several respects from receivers designed for long-distance communications. The receivers used in long-haul communications systems tend to be optimized for speed, sensitivity, and dynamic range at the expense of simplicity and low power consumption; they typically incorporate large electronic circuits and are often built on more than one chip, with several discrete components. Receivers for local optical interconnections, on the other hand, must be simple, low-power circuits which occupy little area. They must also be capable of operating while closely spaced in arrays with VLSI circuits in close proximity, and the signal path must be DC-coupled in order to allow the transmission of unformatted signals. The receivers for local optical interconnections do not need to have as high a bandwidth as those designed for long-distance communications, because each channel needs to carry much less data. For example, while the typical processor today runs at a clock speed of greater than 1 GHz, the off-chip signals generally run at speeds of an order

of magnitude less. Computer processor performance and efficiency can be greatly increased by increasing the off-chip data rates to more closely match the on-chip data rates

At the most basic level, an optical receiver consists of an optical detector fitted to an electronic amplifier, and amplifiers for optical receivers are well developed and mature. The most important characteristics of an optical communications receiver are maximum allowable bit rate, followed by small latency, the time delay between optical input and digital output. By optimizing the receiver stages, bit rate can be maximized and latency can be kept to a minimum. While it is desirable to operate these receivers at low levels of incident optical power, sensitivity is of much less importance than it is in a long-distance communications receiver, as is the need for large dynamic range. Requirements for speed, latency, sensitivity, and dynamic range must be considered with a strong emphasis on simplicity, low power consumption, and low chip area.

5.2 Receiver Fabrication Technology

The technology that is proposed for use in this work is Nortel's BATMOS [5.1], a BiCMOS technology. BATMOS is a 0.8 μm , triple-level-metal silicon BiCMOS technology which is capable of operation in both analog and digital modes. BiCMOS is the name for those technologies that merge CMOS and bipolar technologies together, thereby enabling the strengths of one to compensate for the weaknesses of the other [5.2]. The strengths of CMOS are its low DC power dissipation and its high packing density, making it appropriate for use in producing VLSI circuits, while its weakness is the limited

ability of CMOS to drive capacitive loading which limits the speed of its circuits. The strengths of bipolar technology, on the other hand, are the high speeds achievable by circuits fabricated with this technology; unfortunately, however, the power consumption of bipolar circuits is higher than that of CMOS. Possible applications of optical interconnections lie in interconnecting high-speed processors. For these applications, a BiCMOS technology is ideally suited to provide both bipolar circuitry for the receiver and CMOS technology for the VLSI circuits which processors require.

5.3 Receiver Sensitivity

The receiver in an optical interconnection system is comprised of a photodetector and an amplifier. The purpose of the detector is to detect the incident optical signal and convert it into an electrical current, while the purpose of the amplifier is to convert this electrical current into a usable electrical signal. It is important that the receiver introduces as little noise to the signal as possible so that the receiver has the best possible sensitivity. The sensitivity is defined as the amount of optical power required to reach the receiver in order to achieve a given *bit error rate* (BER) in a digital link [5.3]. Sensitivity is often given as the average optical power required for a BER of 10^{-15} . The receiver sensitivity is a function of both the photodetector and its associated amplification circuitry, and both must perform well if the receiver is to have optimum sensitivity.

5.3.1 Bit Error Rate

In a digital system, the optical intensity has a limited number of discrete values. The binary digital system is by far the most common, and is characterized by the presence of two intensity states: zero intensity which represents the binary 0 and some other, nonzero intensity representing the binary 1. If we assume that the probability of transmitting a binary 1 is equal to that of a binary 0, a detection threshold could be set at the midpoint between these two binary levels. Because noise is a random parameter however, there exists a finite probability that noise will cause an intended binary 0 to be received as a binary 1 or vice versa, thus producing a bit error. The probability of the occurrence of a bit error is termed the bit error rate. If it is assumed that the probability of transmitting a binary 1 and 0 are equal, that there is no statistical correlation between successive bits at the transmitter, and that no correlation between noise in successive bit periods exists, then nearly everything that is needed to be known about the signal path is contained in the bit error rate (BER). The BER is, therefore, an important system parameter.

5.3.2 Noise Sources

A photodetector must perform well in several areas. It must have a high quantum efficiency, low dark current, fast response time, and low capacitance. For a receiver to have good sensitivity however, its detector need only have high quantum efficiency, low dark current, and low capacitance - sensitivity being unaffected by response time. As previously mentioned, the MSM detector has an inherently low capacitance and its dark

current and quantum efficiency can be made respectively low and high by choice of material composition. The receiver amplifier however, must also have good noise performance in order for the receiver to have high sensitivity.

Noise introduced by the receiver is grouped into different classifications; they are shot noise, thermal noise, dark current, and amplifier noise. A photocurrent is the sum of many discrete events which include electron-hole pair production and the subsequent drift of these charge carriers due to the local electric field. Since each electron-hole pair produced gives rise to a current pulse, the total photocurrent is the sum of many current pulses. As such, the total photocurrent is not smooth and continuous, but rather fluctuates about an average value. The fluctuations are called shot noise, and its mean-squared value is given by:

$$\langle i_{sh}^2 \rangle = 2qIB \quad (5-1)$$

where q is the electronic charge, I is the average current, and B is the system bandwidth. Since shot noise is dependant on the average current level present, and there are two distinct current states in a binary transmission system, there must exist two levels of shot noise: one for the case where a photocurrent is present, and the other for when no photocurrent is present. Always, the average current is the sum of photocurrent and dark current (that current produced when zero optical signal is present). When the binary state is 'no optical signal present' however, only the average dark current determines the shot noise, since the photocurrent in this case is zero.

Thermal noise (also called Johnson noise) describes the random motion that electrons exhibit when in a resistor, and as one might expect the intensity of this motion is directly proportional to temperature. The effect of such random motion is to produce a

randomly varying voltage or current. The mean-squared values of thermal voltage and current noise are:

$$\langle v_{th}^2 \rangle = 4kTRB \quad (5-2)$$

$$\langle i_{th}^2 \rangle = \frac{4kTB}{R} \quad (5-3)$$

where k is Boltzmann's constant and R is the value of the resistor in ohms.

The final category of noise is amplifier noise, that is the noise introduced by the receiver front end or preamplifier. The preamplifier provides the first amplification of the signal following the photodetector. It is here that the detected signal is weakest, and as such is most vulnerable to contamination by noise. Noise from the following amplifier stages will be assumed negligible here. The amount of noise produced depends on the structure of the preamplifier, and whether the transistor in the preamplifier is a bipolar junction transistor (BJT) or a field effect transistor (FET). Whatever the structure of the preamplifier though, the noise produced in the amplifier will be a combination of shot noise and thermal noise in various parts of the amplifier. Moreover, following the treatment of Smith and Personick [4.1], the noise from the amplifier can be described as the sum contributions from a signal noise current generator, a shunt noise current generator, and a series noise voltage generator. Figure 5-1 illustrates an amplifier with its associated noise sources referred to the input. In this equivalent circuit C_d and C_s represent the detector and stray capacitance, while R is a load or feedback resistance. A_{pre} represents an ideal, noise-free preamplifier with input impedance Z_a , and shunt and series noise sources i_a and e_a , respectively, while the noise generator i_d represents photodiode dark current. It is convenient to refer all noise due to the preamplifier circuit to the input

Figure 5-1: Noise sources of a receiver front end.

of the receiver. In the case of a FET-based transimpedance amplifier, the input-referred noise current is expressed using conventional analysis [4.1] as:

$$\langle i_{in}^2 \rangle = 2q(I_{gate} + I_{dark})\Delta f + \frac{4kT}{R_f}\Delta f + 4kT\Gamma \frac{(\omega C_T)^2}{g_m}\Delta f \quad (5-4)$$

where the first term represents the shot noise due to the transistor gate leakage current and detector dark current, respectively; the second term is thermal noise in the feedback resistor R_f ; and the third term gives the contribution due to thermal noise in the conducting channel of the transistor. In addition to shot noise due to gate leakage, the FET is also a source of thermal noise from the drain of the FET. Thermal noise in the FET is given as $4kT\Gamma/g_m$. The noise figure of the FET is given by Γ ($\Gamma \approx 0.7$ in Si and $\Gamma \approx 1.1$ for GaAs), and when referred to the input the channel noise is given as the third term of equation (5-4). C_T is the total input capacitance including C_d , and C_s , and Δf is an infinitesimally small bandwidth.

After all of the noise sources have been referred to the input, the noise voltage at the output can be determined by evaluating the integral, over all frequencies, of the product of the input noise current spectral density and the transfer function Z_a of the

preamplifier squared. The spectral density of the input noise is simply equation (5-4) divided by Δf . The resulting output-referred noise voltage is then:

$$\langle v_{out}^2 \rangle = \int_0^\infty \left[2q(I_{gate} + I_{dark}) + \frac{4kT}{R_f} \right] |Z_a(f)|^2 df + \int_0^\infty 4kT \frac{(\omega C_T)^2}{g_m} |Z_a(f)|^2 df \quad (5-5)$$

The total circuit noise is then referred to the input by dividing the output-referred noise voltage by the squared magnitude of the transfer function at $f=0$:

$$\langle i_c^2 \rangle = \left[2q(I_{gate} + I_{dark}) + \frac{4kT}{R_f} \right] \int_0^\infty \frac{|Z_a(f)|^2}{|Z_a(0)|^2} df + 4kT \frac{(2\pi C_T)^2}{g_m} \int_0^\infty f^2 \frac{|Z_a(f)|^2}{|Z_a(0)|^2} df \quad (5-6)$$

In order to evaluate the amplifier noise, the preamplifier transfer function Z_a must be determined. The following analysis follows that of Morikuni et al [5.4]. A general transimpedance topology, like the one illustrated in Figure 5-2, is considered. By examining the small-signal equivalent circuit of a FET-based transimpedance amplifier given in Figure 5-3, A'_v , the open-loop voltage gain modified to take into account the effects of feedback loading can be determined using an amplifier internal current source of value $g_m v_i$ and an internal load resistance R_l :

$$A'_v = \frac{-g_m + \frac{1}{R_f} + j\omega C_f}{\frac{1}{R_l} + \frac{1}{R_f} + j\omega(C_o + C_f)} \quad (5-7)$$

where R_l and C_o are the load resistance and capacitance of the circuit, and R_f and C_f are the feedback resistance and capacitance, respectively, and $g_m = (\mu_s C_{ox})(W/L)(V_{GS} - V_t)$ is the MOSFET transconductance, where μ_s is the carrier mobility, C_{ox} is the per unit area capacitance of the gate, W/L is the width to length ratio of the transistor, V_{GS} is the gate bias voltage, and V_t is the threshold voltage of the MOSFET. When the transimpedance

Figure 5-2: General transimpedance amplifier topology.

amplifier uses MOSFET devices, the load resistors R_f and R_l are assumed to be active loads since these generally offer better performance in MOS circuits than do passive feedback resistors. The parasitic capacitance of C_f , C_o , and C_{in} are the sum total feedback, output and input capacitance values, respectively. They are expressed as:

$$C_f = C_{gd} + C_{dsf} \quad (5-8)$$

$$C_o = C_{ds} + C_{gdf} + C_{dsl} + C_{gdl} + C_{next} \quad (5-9)$$

$$C_{in} = C_{pd} + C_{gs} + C_{gsf} + C_{pad} \quad (5-10)$$

where the subscripts d , g , and s refer to the drain, gate and source, respectively, of a transistor. The subscript f refers to the feedback transistor; the subscript l refers to the load transistor; C_{pd} is the photodiode capacitance; C_{pad} is the total pad capacitance from pads linking the photodiode to the preamplifier; and C_{next} refers to the input capacitance of whatever stage follows the preamplifier. The transimpedance transfer function can then be expressed as:

$$Z_a(\omega) = \frac{v_{out}(\omega)}{i_{in}(\omega)} = \frac{R_f \left(\frac{A'_v}{1 + A'_v} \right)}{1 + j\omega R_f \left(C_f - \frac{C_{in}}{1 + A'_v} \right)} \quad (5-11)$$

$$= \frac{A + j\omega B}{C + j\omega D + (j\omega)^2 E}$$

where the constants $A - E$ are:

$$A = R_o(1 - g_m R_f) \quad (5-12)$$

$$B = R_l R_f C_f \quad (5-13)$$

$$C = 1 + g_m R_l \quad (5-14)$$

$$D = R_l(C_{in} + C_{out}) + R_f(C_f + C_{in}) + g_m R_l R_f C_f \quad (5-15)$$

$$E = R_l R_f [(C_{in} + C_{out})C_f + C_{in}C_{out}] \quad (5-16)$$

Knowing now the transfer function, it is possible to solve the integrals of equation (5-6); they have been evaluated [5.4] as:

$$\int_0^\infty \frac{|Z_a(f)|^2}{|Z_a(0)|^2} df = \frac{C}{4D} \quad (5-17)$$

$$\int_0^\infty f^2 \frac{|Z_a(f)|^2}{|Z_a(0)|^2} df = \frac{C^2}{16\pi^2 DE} \quad (5-18)$$

Figure 5-3: Small-signal equivalent circuit of transimpedance amplifier.

where $C - E$ are given above. By substituting equations (5-17) and (5-18) into equation (5-6) the total input circuit noise for transimpedance preamplifiers whose transfer function matches that given by equation (5-11) is given as:

$$\langle i_c^2 \rangle = \left[2q(I_{gate} + I_{dark}) + \frac{4kT}{R_f} \right] \left(\frac{C}{4D} \right) + 4kT \frac{(2\pi C_{in})^2}{g_m} \left(\frac{C^2}{16\pi^2 DE} \right) \quad (5-19)$$

For the case in which the transimpedance preamplifier is based on a bipolar junction transistor (BJT), a similar analysis ensues. When the small-signal circuit model is equivalent to that given in Figure 5-4, the open-loop voltage gain modified to take into account the effects of feedback loading is expressed exactly as it is in equation (5-19) with the exception that the coefficients g_m , C_{in} , C_f , and C_{out} have, in this case, different values than in the MOSFET case, and I_{gate} is replaced with I_{base} for bipolar transistors. Here the transconductance is given by $g_m = qI_E/kT$ where I_E is the emitter current of the BJT. The resistor r_π is the dynamic resistance of the base-emitter junction and is given as $r_\pi = \beta_o/g_m$ where β_o is the small-signal forward current gain of the transistor. The BJT

Figure 5-4: Small-signal circuit model of BJT-based transimpedance amplifier.

parasitic capacitance elements are given as:

$$C_{in} = C_{pd} + C_{pad} + C_{\pi} \quad (5-20)$$

$$C_{out} = C_{next} \quad (5-21)$$

$$C_f = C_c \quad (5-22)$$

where C_{π} represents the total effective capacitance associated with the base-emitter junction of the transistor, and C_c is the collector-base junction capacitance.

Now that the circuit noise has been determined, the signal noise can be analyzed. It is necessary to know both circuit and signal noise in order to determine the signal-to-noise ratio. The average output voltage produced by an incident optical signal $P(t)$ incident on an MSM-type detector can be expressed as [5.5]:

$$\langle v_{out}(t) \rangle = \int_{-\infty}^{\infty} \Re P(\tau) Z_a(t - \tau) d\tau \quad (5-23)$$

where $\Re$ is the detector responsivity as previously defined, and $Z_a(t)$ is the preamplifier impulse response (the Fourier transform of $Z_a(\omega)$). The input-referred current noise due to this voltage fluctuation is given as:

$$\langle i_s^2(t) \rangle = q \int_{-\infty}^{\infty} \Re P(\tau) H_a^2(t - \tau) d\tau \quad (5-24)$$

where $H_a(t)$ is given by:

$$H_a(t) = \frac{Z_a(t)}{Z_a(0)} \quad (5-25)$$

by rewriting the expression for $Z_a(\omega)$ given previously, $H_a(\omega)$ can be written as:

$$H_a(\omega) = \frac{Z_a(\omega)}{Z_a(0)} = \left(\frac{C}{E\omega_o} \right) \left(\frac{\omega_o}{\left(\frac{D}{2E} + j\omega_o \right)^2 + \omega_o^2} \right) \quad (5-26)$$

where ω_o is defined as:

$$\omega_o = \sqrt{\frac{C}{E} - \left(\frac{D}{2E}\right)^2} \quad (5-27)$$

and $C - E$ are as given previously. The above equation (5-27) is the Fourier transform of:

$$H_a(t) = \left(\frac{C}{E\omega_o}\right) e^{\frac{Dt}{2E}} \sin(\omega_o t) u(t) \quad (5-28)$$

where $u(t)$ is the unit step function; equation (5-28) represents the impulse response of the receiver and can be directly inserted into equation (5-24). Signal noise is a maximum at $t = 0$, which is the point of maximum signal and is therefore the most useful point at which to determine signal noise. When the incident optical signal $P(t)$ is a square pulse such that $P(t) = P_o$ for $-T/2 < t < T/2$ and 0 otherwise, the convolution integral of the input-referred signal current noise of equation (5-24) can be expressed as [5.4]:

$$\langle i_s^2(0) \rangle = \frac{q\mathcal{R}P_o}{4} \left(\frac{C}{E\omega_o}\right)^2 \cdot \left[e^{\frac{DT}{2E}} \left(\frac{\frac{D}{2E} \cos(\omega_o T) - \omega_o \sin(\omega_o T)}{\left(\frac{D}{2E}\right)^2 + \omega_o^2} - \frac{2E}{D} \right) - \left(\frac{\frac{D}{2E}}{\left(\frac{D}{2E}\right)^2 + \omega_o^2} - \frac{2E}{D} \right) \right] \quad (5-29)$$

where T is the signal period and T^{-1} is the signal bit rate. Note that while circuit-induced noise is independent of signal bit rate, signal noise is not. Total noise is expressed as the sum of circuit and signal noise:

$$\langle i_{noise}^2 \rangle = \langle i_c^2 \rangle + \langle i_s^2 \rangle \quad (5-30)$$

It should be noted, however, that when equations (5-19) and (5-29) are evaluated and compared, signal noise is usually much less than circuit noise.

The signal-to-noise ratio (SNR) can now be determined by using the signal term expressed as:

$$i_{sig} = \frac{\eta q P \lambda}{hc} \quad (5-31)$$

The power SNR term is then given by:

$$SNR = \frac{i_{sig}^2}{\langle i_{noise}^2 \rangle} \quad (5-32)$$

The signal and noise levels determine the receiver bit error rate. The bits that are actually transmitted originate at an information source that transmits a stream of bits. In a binary communication system, two types of bits are transmitted: one for $x = 0$, and one for $x = 1$. The probability that the transmitted x lies somewhere between x and $x + dx$ is $P(x)dx$, where $P(x)$ is the probability density. The receiver in the system quantizes the detected bits into either $y = 0$ or $y = 1$, and the probability density is $P(y)$. The probability that y is received *given that* x is transmitted $P(y|x)$ is termed the conditional probability density, and has a peak value at m_x . Figure 5-5 gives an example of the conditional probability densities $P(y|0)$ and $P(y|1)$. Since the two functions $P(y|0)$ and $P(y|1)$ overlap each other, a threshold level must be established in order to separate the two. All bits with a y value less than threshold will be declared as “0”, and all those with a y value greater than threshold will be declared as “1”. This strategy will then cause all $x = 0$ bits with a received value greater than threshold to be misinterpreted as “1” bits, and all $x = 1$ bits with a received value less than threshold to be mistakenly decoded as “0” bits.

Figure 5-5: The conditional probability density distributions $P(y|x)$ for the cases in which the transmitted bit x is a “0” and a “1” respectively. The variable y is the received value of the signal bit.

The conditional probability density can be approximated by a gaussian waveform in which:

$$P(y|x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\left(\frac{y-m_x}{\sqrt{2}\sigma_x}\right)^2\right] \quad (5-33)$$

where σ_x is the standard deviation of the distribution, and m_x is the mean; each of these have different values for $x = 0$ and $x = 1$. The area under a gaussian curve and to one side of a point $z = \alpha$ can be calculated by integrating the gaussian function over the appropriate limits and coming up with the well-known Q -function:

$$Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz \approx \frac{1}{\sqrt{2\pi}\alpha} \exp\left(-\frac{\alpha^2}{2}\right) \quad (5-34)$$

where the above approximation is accurate to within one percent at $\alpha = 3$; the accuracy improves at higher values of α . The probability of error of the system may be determined by multiplying the probability of transmitting a “1” by the probability of mistaking a “1”

for a “0” and adding this to the product of the probability of transmitting a “0” and the probability of mistaking a “0” for a “1”.

The lower limit for error probability occurs for the case where the threshold is set so that the tails of each curve contain equal area, assuming that a “0” and a “1” are equally probable. In this case the total probability of error reduces to the probability of mistaking a “1” for a “0” or that of mistaking a “0” for a “1”; either will do since they are the same in this case. By setting the threshold θ where the height of the gaussian curves are equal such that:

$$Q(z) = Q\left(\frac{\theta - m_0}{\sigma_0}\right) = Q\left(\frac{m_1 - \theta}{\sigma_1}\right) \quad (5-35)$$

by equating the above arguments, the threshold θ can be determined:

$$\theta = \frac{m_0\sigma_1 + m_1\sigma_0}{\sigma_0 + \sigma_1} \quad (5-36)$$

Substituting θ back in, the probability of error can be expressed:

$$BER = Q\left(\frac{m_1 - m_0}{\sigma_0 + \sigma_1}\right) = Q(\gamma) \quad (5-37)$$

where the argument γ in equation (5-37) can be described as the voltage signal-to-noise ratio (SNR). While strictly speaking, bit error rate and probability of error are not synonymous it is common practice to regard them as so, and they will be treated as synonyms here. The Q -function for the noise functions of Figure 5-5 is given in Figure 5-6. Note that for a BER of 10^{-9} γ is approximately equal to 6, and for a BER of 10^{-15} γ is approximately 8. Receiver sensitivity is defined as the minimum received signal power that allows the receiver to produce an output signal that meets the required bit error rate.

By combining equations (5-31), (5-32) and (5-37), and assuming that signal noise is much less than circuit noise and that optical power in the digital '0' state is zero, the sensitivity of a receiver can be expressed as [5.4]:

$$\eta \bar{P} = \frac{hc}{\lambda q} \gamma \sqrt{\langle i_c^2 \rangle} \quad (5-38)$$

When optical power in the digital '0' state is not zero, then the receiver sensitivity is given as [5.4]:

$$\eta \bar{P} = \frac{hc}{\lambda q} \cdot \frac{1+r}{1-r} \cdot \gamma \sqrt{\langle i_c^2 \rangle} \quad (5-39)$$

where r is the ratio of optical power in the low state to that in the high state.

The MSM detector, as stated previously, has a much lower capacitance per unit active area than that of a $p-i-n$ detector - approximately 1/3 the capacitance of a $p-i-n$. This will aid greatly in improving the sensitivity of a receiver unit. More than this however, integrating the detector onto the same chip as the amplification circuitry will eliminate all of the parasitic capacitance associated with interconnecting discrete elements; thus further reducing the input capacitance. Another benefit of the monolithic integration of photodetector and amplifier is the reduction of the length of the coupling connection between detector and amplifier. The shorter coupling wire reduces the inductance of the connection as well as the switching noise picked up inductively from nearby circuits [1.26]. Yet another advantage is increased ease of manufacture (once the IC fabrication process to be used has been modified to include the photodiode).

Figure 5-6: The Q -function for determining bit error rate for the case where noise amplitude resembles a gaussian distribution.

5.4 Wideband Amplifiers

When using monolithic devices to build wideband amplifiers, the limitations of integrated devices impose restrictions on possible designs. Most notably, large decoupling capacitors are unavailable to the monolithic circuit designer, and so most monolithic circuits are DC coupled. However, the same absence of large capacitors also aids the monolithic designer. Parasitic and stray capacitance which shunt the signal path are minimized in a monolithic circuit, making it possible to design and build high-performance wideband monolithic amplifiers.

The availability of well-matched, high-frequency transistors is the monolithic circuit designer's greatest asset. The *npn* bipolar transistor has the best high-frequency characteristics among integrated devices. For that reason, *npn* devices are used in the signal path of nearly every monolithic wideband amplifier. If used at all, *pnp* devices are used in biasing applications. The loads which are used in wideband designs are resistive, as opposed to the active loads used in low-frequency amplifiers. Resistive loads have much better frequency and noise characteristics than do active loads.

5.4.1 High-Frequency Transistors

In order to design a wideband amplifier having a moderate gain of 20-30 dB, transistors having a unity current gain-bandwidth product (f_T) of 5-10 times that of the desired 3 dB bandwidth of the amplifier are required as a general rule [5.6]. To create transistors that have good high-frequency characteristics and a high f_T , several design considerations must be observed. Firstly, all parasitic capacitance elements, especially those between the transistor's base and collector, as well as resistance elements must be minimized. Secondly, the transistor base width must be minimized to reduce the transit time of minority carriers through the base. Finally, the transistors must be operated at sufficiently high emitter current densities, where f_T is nearly maximal. To meet these requirements, junction areas must be minimized while at the same time not allowing series resistance to increase; also emitter efficiency must remain high at relatively high emitter currents. A stripe geometry for emitter, base and collector contacts best meets the above requirements. Figure 5-7 illustrates a typical device layout for a stripe geometry

transistor. The emitter area is comprised of long and narrow n^+ regions resembling stripes. The stripe geometry allows a maximum emitter periphery: area ratio and therefore avoids current-crowding effects, while at the same time this geometry minimizes junction capacitance. Base contacts are placed adjacent to emitter contacts, also in a stripe geometry. The close proximity of stripe geometry base and emitter contacts allows for low base-spreading resistance and low base-emitter capacitance. The collector contact surrounds the base region in an effort to minimize the collector series resistance, and its close proximity to the base contacts creates a low collector-base junction capacitance. The minimum width of these stripe contacts, as well as their minimum separation, is determined by the minimum feature size of the technology used to fabricate the transistors. Obviously, the minimum feature size that a technology can produce greatly affects the high-frequency performance of stripe-geometry transistors.

Figure 5-7: Typical layout of a high-frequency *npn* transistor. The shaded regions indicate contact windows.

5.4.2 Amplifier Designs

By far, the most commonly used gain block in analog circuit design is the common-emitter gain stage shown in Figure 5-8; this gain block is capable of providing both voltage and current amplification. However, the gain-bandwidth product of the common-emitter amplifier is not large, due to the Miller effect. The Miller effect states that a capacitance existing between the base and collector of a transistor is equivalent to a capacitance, applied between the transistor base and ground, which has a value equal to that of the true capacitance multiplied by the voltage gain of the amplifier. This large effective capacitance on the input of the amplifier increases the RC time constant and lowers the speed of the amplifier. The Miller effect can be explained with the help of the hybrid- π high-frequency device model for a bipolar transistor. Figure 5-9 gives the hybrid- π model for a bipolar transistor, along with the parasitics r_b , r_{cs} , and C_c inherent to bipolar transistors. The component r_o is the output resistance due to the Early effect, and C_{cs} is the collector-substrate junction capacitance. The transistor parameters r_π and C_π are given as:

$$r_\pi = \frac{\beta_o}{g_m} \quad (5-40)$$

Figure 5-8: Basic common-emitter gain stage.

$$C_{\pi} = \frac{g_m}{\omega_T} - C_c \approx \frac{g_m}{\omega_T} \quad (5-41)$$

where ω_T is the unity gain bandwidth of the transistor. The hybrid- π model of Figure 5-9 can be simplified, using the Miller effect, into the approximation illustrated in Figure 5-10. In this approximation:

$$C_t = C_{\pi} + C_c(1 + g_m R_L) \quad (5-42)$$

however, when $g_m \gg 1$,

$$C_t = C_{\pi} + C_c g_m R_L \quad (5-43)$$

The simplified model shown in Figure 5-10 illustrates the Miller-effect approximation. It should be noted that the Miller approximation assumes that the circuit load is purely resistive, and that the capacitive loading effects at the output are negligible.

The frequency response of the common-emitter stage can be approximated using the Miller approximation hybrid- π transistor model of Figure 5-10:

$$A_v(s) = \frac{v_{out}}{v_{in}} = \frac{A_{vo}}{1 - s / p_1} \quad (5-44)$$

Figure 5-9: Hybrid- π model of a bipolar transistor.

Figure 5-10: Miller approximation of hybrid- π model.

where $s = j\omega$ is the complex frequency, p_1 is the value of the dominant pole, and A_{vo} is the low-frequency gain of the circuit. The low-frequency gain can be determined by examining the circuit of Figure 5-10:

$$A_{vo} = -g_m R_L \frac{r_\pi}{R_S + r_b + r_\pi} \approx -g_m R_L \quad (5-45)$$

and the dominant pole is expressed as:

$$p_1 = -\frac{R_S + r_b + r_\pi}{(R_S + r_b)r_\pi} \cdot \frac{1}{C_\pi + C_c g_m R_L} \quad (5-46)$$

The -3 dB bandwidth is simply the magnitude of p_1 for a single-pole response.

In order to provide sufficient gain for the amplifier of an optical receiver, several stages must be cascaded together; and when several common-emitter stages are joined together to form a cascade, one stage's R_L becomes the next stage's R_S as illustrated in Figure 5-11. As equation (5-46) shows, R_L (and hence R_S) must be small in order to have a high bandwidth. When R_L is small however, the gain of the stage is low as equation (5-45) shows. In this way it is possible to exchange amplifier gain for bandwidth. Ideally, the gain-bandwidth product of the transistor model of Figure 5-10 is [1.30]:

$$GB = \frac{g_m}{C_i} \quad (5-47)$$

However, when a cascade of common-emitter stages form an amplifier, the ideal gain-bandwidth product of each stage cannot be reached because of conflicting requirements

Figure 5-11: Circuit diagram of amplifier comprised of cascaded common-emitter stages.

on the collector resistance R_c of each stage. In order to have the highest possible bandwidth, the input of a common-emitter stage requires a low impedance source (R_c for the previous stage) in order to charge the capacitance C_i quickly. At the same time, in order to rapidly discharge the C_i of the following stage it is necessary for the load impedance (R_c of the stage in question) to be large. This conflicting requirement on R_c means that a compromise must be reached on the value of R_c , and as such the gain-bandwidth product of the resulting amplifier is much less than the product of the gain-bandwidth products of the individual stages.

To achieve the highest possible gain-bandwidth product for a cascade of individual amplifier stages, it is advantageous to alternate between stages which have requirements for high and low source impedance, respectively. By employing local

feedback, stages having the necessary impedance requirements can be created. Shunt feedback is used to form the gain stage of Figure 5-12. This stage is a *transimpedance* amplifier and provides a voltage output from a current input. The shunt-feedback stage has low input and output impedance levels. The low-frequency transimpedance gain of the circuit can be obtained from the small-signal equivalent circuit of Figure 5-12 (b) [5.6]:

$$A = \frac{v_{out}}{i_{in}} = \frac{\frac{1}{g_m} - R_f}{1 + \frac{(R_L + R_F + r_\pi)}{\beta_o R_L}} \quad (5-48)$$

and since generally $\beta_o \gg 1$ and $g_m \gg 1$, the above expression can be simplified to

$$A \approx -R_F \quad (5-49)$$

The input resistance may be determined by similar examination of the small-signal equivalent circuit to be:

$$R_{in} = \frac{r_\pi (R_L + R_F)}{r_\pi + R_L + R_F + \beta_o R_L} \quad (5-50)$$

Figure 5-12: Single-stage shunt-feedback amplifier: (a) circuit schematic; (b) small-signal model circuit.

Whereas the output resistance can be calculated as:

$$R_{out} = \frac{r_{\pi} + R_f}{1 + \frac{R_f + r_{\pi}}{R_L} + \beta_o} \approx \frac{r_{\pi} + R_f}{\beta_o} \quad (5-51)$$

As such, both the input and output resistance of the shunt feedback stage are relatively low.

The local series-feedback gain stage illustrated in Figure 5-13 has, in some ways, complete opposite performance characteristics of those of the shunt-feedback gain stage. Such an amplifier has relatively high input and output impedance levels and provides a current output from a voltage input; it is often called a *transadmittance* amplifier. By using the small-signal circuit model of Figure 5-13(b), the transadmittance gain of the circuit can be calculated as:

$$A' = \frac{i_{out}}{v_{in}} = \frac{-1}{R_E} \cdot \frac{1}{1 + \frac{1}{R_E} \left(\frac{1}{g_m} + \frac{R_E}{\beta_o} + \frac{r_b}{\beta_o} \right)} \quad (5-52)$$

Figure 5-13: Single-stage series-feedback amplifier: a) circuit schematic; b) small-signal model circuit.

and since usually $\beta_o \gg 1$, and $g_m R_E \gg 1$, equation (5-52) above can be simplified to:

$$A' \approx \frac{-1}{R_E} \quad (5-53)$$

The input resistance of the circuit of Figure 5-13 b) can be similarly calculated as:

$$R_{in} = r_b + r_{\pi} + R_E(\beta_o + 1) \approx \beta_o R_E \quad (5-54)$$

and the output resistance can also be determined:

$$R_{out} \approx r_o \left(1 + \frac{\beta_o g_m R_E}{\beta_o + g_m R_E} \right) \quad (5-55)$$

By alternately cascading transimpedance gain stages with transadmittance gain stages, it is possible to satisfy the impedance requirements of both. The transimpedance stage can provide a low-impedance voltage signal for the following transadmittance stage, which in turn can provide the high-impedance drive required by the transimpedance stage following it. When one transimpedance stage follows one transadmittance stage, the resulting voltage gain can be approximately obtained by multiplying the gains of the individual stages:

$$A_{vo} = \frac{v_{out}}{v_{in}} = \left(\frac{i_{o1}}{v_{in}} \right) \left(\frac{v_{out}}{i_{o1}} \right) \approx \frac{R_F}{R_E} \quad (5-56)$$

Likewise for the case where one transadmittance stage follows one transimpedance stage:

$$A_{io} = \frac{i_{out}}{i_{in}} = \left(\frac{v_{o1}}{i_{in}} \right) \left(\frac{i_{out}}{v_{o1}} \right) \approx \frac{R_F}{R_E} \quad (5-57)$$

5.5 Integrated Receiver Design

The design philosophy for the optical interconnection receiver contradicts that of the traditional, long-haul fibre-optic telecommunications receiver. Long-haul systems have fiber attenuation that ranges from 0 to 20 dB, while short-haul links have consistently very nearly zero attenuation between the transmitter and the receiver. The optical interconnection receiver is a short-haul receiver and has a relatively high-power optical signal incident on it, which receiver photodetectors convert into large current signals. By eliminating the up to 20 dB of signal attenuation that are commonly found in long-haul systems, the gain and dynamic range requirements on the short-haul receiver amplifier are thereby relaxed by 20 dB. And since optical interconnections are able to act in parallel, the bit rate requirement on the receiver is relaxed. The performance goal for the intraboard communication link receiver was to have 16 amplifiers on one die, each having an operating speed of 622 Mbit/s, while amplifying the signal to ECL levels (800 mV_{p-p}), and minimizing crosstalk between amplifiers. Other desirable quantities are: a sensitivity of -20 dBm, a dynamic range of 10 dB on the input, an internal delay of less than 1 ns, and a power consumption of less than 5 mW per channel. It was also desired that the amplifier was to be coupled to an output stage capable of driving 15 pF loads, as this was the stray capacitance estimated to exist when driving a chip on the same board but in a separate package.

The integrated receiver for local area network communication might consist of a MSM photodiode (see Chapter 3) or some other photodetector followed by a transimpedance preamplifier which converts current signals from the MSM diode into

voltage signals. The voltage signals from the preamplifier would then be fed into an amplifier circuit, which acts as a comparator and has full ECL voltage swing capability. The structure of this type of receiver is illustrated in Figure 5-14.

Figure 5-14: Structure of the integrated receiver.

In order to maximize the sensitivity of the receiver, as expressed in equation (5-38), the circuit noise of the preamplifier must be minimized. Preamplifier circuit noise is given in equation (5-19). By examining equations (5-19) and (5-11) - (5-16), it may be discovered that there are many ways to reduce noise for specific circuits. However, for a general transimpedance stage, circuit noise can be reduced only by increasing the sizes of R_f , C_f , or C_o , and by decreasing the size of the total input capacitance C_{in} . However, equation (5-11) also states that in order for a transimpedance amplifier to have a large bandwidth, the coefficient C must be proportionally large or the coefficient E must be inversely proportional to bandwidth. This means that for high-speed operation, the values of the load resistor R_l and g_m must be relatively high whereas the values of R_f , C_f , C_o and C_{in} must be relatively small. Therefore the only way to increase sensitivity without sacrificing high-speed operation is to minimize the size of the total input capacitance C_{in} . The advantage of using a MSM detector is the previously mentioned lower capacitance of this type of detector over an equivalent-size *p-i-n* detector. Moreover, by integrating the

detector on-chip with the preamplifier, pad and excessive trace capacitance is eliminated. Input capacitance and thereby circuit noise is kept to a minimum, and receiver sensitivity is improved.

The receiver preamplifier design for a local area network should be a simple transimpedance stage wherein the feedback resistance R_f is in the range of 300 - 1000 Ω . Below that range amplifier gain tends to be too low, whereas above the range crosstalk and instability are risked. The feedback supplied by R_f should be limited to a single stage [5.6]. The main advantages of this design over the more common multi-stage feedback loop are the former's higher bandwidth, flatter frequency characteristics of both magnitude of the transimpedance and of phase delay, lower sensitivity of the frequency characteristics to design and process variations, and fewer stability problems. If the preamplifier and main amplifier are, unlike conventional designs, integrated onto the same chip there is no need to provide output impedance matching with a source follower stage after the preamplifier. This power-saving step can save 10 - 20 mW per channel, depending on the technology used. The receiver preamplifier design for local area networks is illustrated in Figure 5-15.

The main amplifier design for local communications applications abides by circuit principles which are well suited for wide-band applications. The stages employed in the amplifier design all contain some form of feedback. Mismatching between succeeding stages could be employed as explained previously so that, for example, a high input impedance stage is fed by a low output impedance stage. This principle is used in the differential amplifier stage employed in the main amplifier design. Two types of main amplifier were considered: automatic gain control (AGC) amplifiers and limiting

Figure 5-15: Circuit diagram of a transimpedance preamplifier together with a MSM photodiode.

amplifiers. A limiting amplifier must have sufficient gain to amplify the smallest expected input signal to a voltage level that is large enough for the following circuitry, usually a clock and data recovery circuit, to recognize. Input signals that have larger than the minimum voltage swing cause the output of the amplifier to saturate, and thus limit the maximum output signal swing. Limiting amplifiers have the advantages of simpler design, lower power consumption, and less required silicon area than an AGC, therefore a limiting amplifier design was preferred. One design, which has been extensively simulated by the author using Analog Artist and Hspice, is illustrated in Figure 5-16. Notice that this design uses emitter followers as outputs for the amplifier stage. This improves the impedance transformation between successive amplifier stages and also shifts the DC value of the voltage signal, but care must be taken to ensure that gain peaking is not excessive and that the load capacitance does not lead to instability. The design was simulated on Hspice and found to have a -3 dB bandwidth of 3 GHz. Notice that it consists of a differential transadmittance stage cascoded with a differential transimpedance stage.

Figure 5-16: Differential amplifier cell for wide-band main amplifier.

Differential operation is used because of the numerous advantages it enjoys over single-ended operation. Differential operation considerably reduces circuit noise, timing jitter, electromagnetic interference and crosstalk. Differential operation improves signal pulse shape and it allows the use of smaller voltage swings, which thereby leads to higher operating speed [5.7]. On the other hand, single-ended design has the advantages of lower power consumption and simplicity of design. The amplifier stage as shown in Figure 5-16, was designed to work with a -5 V supply, and as such it consumes a total of 55 mW of power. Much less power consumption is needed for an optical interconnection receiver, but even when designed to work at a supply voltage of -3 V this amplifier still consumed over 30 mW of power. Clearly another design is needed.

Another amplifier which was created and thoroughly simulated was much simpler, smaller, and designed to consume much less power than the previous differential amplifier. It can operate with the same MSM detector proposed earlier, but employs a simple transimpedance amplifier circuit constructed from CMOS and bipolar transistors. A schematic illustration of the amplifier circuit is given in Figure 5-17. The circuit is a BiCMOS structure, employing both CMOS and *npn* bipolar transistors in its design. The circuit in Figure 5-17 has been simulated on Hspice. This circuit has been found capable of driving an 80 fF load at a 3dB data rate of over 400 Mbit/s with a 12 dB dynamic range; the circuit's power consumption at this speed is 0.6 mW when $V_{DD} = 3$ V. The receiver is able to drive loads larger than 1 pF at 400 Mbit/s, albeit with a slightly reduced dynamic range. This is a novel circuit design in which the transimpedance stage is based on a previous design [5.8]. The idea behind the feedback transistors seen in Figure 5-17 is to replace a feedback resistor, which would require too much chip area. Providing the gate of the first feedback transistor with a constant positive potential produces an effective channel resistance. The second feedback transistor is connected as a diode, which improves the dynamic range of the receiver. The bipolar output stage is capable of sourcing and sinking relatively large currents at high speed, and is therefore able to drive large capacitive loads. However, with a maximum bit rate of 400 Mbit/s, this design does not have the bandwidth required for OC-12 transmission rates (622 Mbit/s).

Figure 5-17: Circuit schematic of intraboard receiver circuit.

A compromise between the above two designs was developed in order to reach the previously mentioned goals of 622 Mbit/s at acceptable power consumption. The design is a very simple one consisting of a preamplifier integrated with a limiting amplifier, and is implemented with bipolar npn transistors. The amplifier circuit design uses alternating stages of transimpedance and transadmittance configuration and benefits from the impedance mismatch that is inherent to this type of architecture. The amplifier produces an inverted output. A schematic of the circuit is given in Figure 5-18.

The amplifier consists of three stages: an input stage that acts as a preamplifier followed by a limiting amplifier, and finally an output driver stage. The purpose of the last stage is to act as a current amplifier so that the amplifier is capable of driving 15 pF loads. Each stage has separate power and ground connections that are tied together off-chip. The purpose of separate power and ground circuits is to minimize feedback which

Figure 5-18: Circuit schematic of integrated preamplifier/limiting amplifier for optical receiver. This design is based on Cherry and Hooper's principle of alternating stages to achieve an impedance mismatch between successive stages.

can occur through power and ground lines when they are not separated locally. Such feedback can lead to oscillation of the amplifier. The output stage was originally designed as an emitter follower, as is the case for ECL drivers. Simulations of the entire chip, including extracted parasitic capacitance, however revealed an oscillatory behavior of which the emitter follower output stage was discovered to be the cause. An illustration of the hybrid- π model of the emitter follower simplified for high-frequency operation is given in Figure 5-19. The input resistance of the stage can be determined by dividing the input voltage of the stage by its input current. By looking at Figure 5-19, the input voltage v_i can be found:

$$v_i = v_\pi + \left(\frac{1}{j\omega C_L} \right) (g_m v_\pi + j\omega C_\pi v_\pi) \quad (5-58)$$

Similarly, the input current of an emitter-follower is:

$$i_i = j\omega C_\pi v_\pi \quad (5-59)$$

Figure 5-19: High-frequency hybrid-pi model of an emitter-follower stage.

By dividing equation (5-58) by equation (5-59), the input impedance of an emitter-follower is:

$$Z_i = \frac{1}{j\omega C_\pi} + \frac{1}{j\omega C_L} + \frac{-g_m}{\omega^2 C_\pi C_L} \quad (5-60)$$

The last term in equation (5-60) represents a negative resistance, which can provide the gain and feedback necessary for oscillation to occur. The output stage was redesigned in order to eliminate oscillations that the emitter-follower output stage had caused. Because this is a limiting amplifier, the analog to digital conversion is performed by the amplifier which contains a fixed decision threshold (see section 5.3.2). This means that it does not have the ability to adjust the decision threshold in order to achieve a higher BER, a feature that is available only with AGC amplifiers, and one that was sacrificed for circuit simplicity and low chip area. The input impedance of the amplifier is quite low and, using Equation (5-50) is approximately $R_{in} \cong r_\pi/(1+\beta)$.

When laying out the chip, it was considered important that proper measures were to be taken to reduce on-chip interactions, as these deteriorate the performance of the receiver. Capacitive coupling between external signal lines and inductive coupling through power supply lines can also lead to a deterioration of amplifier characteristics [5.9]. To avoid these effects, several precautions were taken. The receiver stages were arranged one after another in a straight line in their order in the signal path. Such an arrangement keeps the most sensitive section (the detector and preamplifier) away from the output stage where signal swings are largest. The preamplifier was surrounded by a metallic ring of first level metal that was grounded to the substrate. This type of guard ring is able to shield sensitive sections from stray substrate currents. Separate ground and supply pads for each of the receiver cells were added to reduce the interactions through these networks. Moreover, the ground pads were elongated in order to allow for parallel bonding, thereby reducing the effective bond wire inductance and providing a lower impedance to ground. These steps should help minimize the effects of crosstalk between receiver stages, as well as interference from external signal lines and power supplies.

The power consumption of the circuit in Hspice simulations was 5 mW for the preamplifier/limiting amplifier and 30 mW for the output driver. Power requirements for the circuit are 3 V for V_{cc} and 0 V for ground. In Hspice simulations, the threshold input current was 10 μ A, above which a digitally high output level was produced and below which the output was low. In simulations, the output high and low states were 360 mV and 200 mV respectively with a 50 Ω load. While a load of 1 M Ω gave a high and low state of 1.3 V and 400 mV respectively. A latency of 1.5 ns was predicted. A total of 16 separate amplifiers were arranged side-by-side on a die measuring 3.44 x 1.45 mm.

Figure 5-20: Microphotograph of 16-channel limiting amplifier. The input pads are in the bottom row, while the output pads make up the top row. Power and ground pads are each on the left and right side of the die. Four elongated ground pads are visible and each is attached to three bond wires.

Each amplifier has one pad for the input signal and one pad for its output; these pads are on opposite sides of the die and are spaced at $167\text{ }\mu\text{m}$ centre-to-centre. A photograph of one of the chips is given in Figure 5-20. Power and ground pads were placed on the two remaining sides of the circuit. In Figure 5-20, four elongated pads are visible; these pads each have three bond wires which are connected directly to the package floor, which is grounded. All 16 ground wires are approximately 2 mm in length, providing a very low-inductance path to ground for the circuit. Bond wire inductance of 1.0 nH/mm was included when the entire chip was simulated in Hspice.

DC measurements of the limiting amplifier chip were performed on a wafer probe station. The positive rail was supplied with +3 V and the negative was grounded. With a high impedance load applied to the output, a voltage of 1.4 V was produced at the outputs in the high state, whereas a voltage of 0.40 V was produced in the low state. With a 50Ω load on the output, a voltage of 0.40 V was produced in the high state, and a voltage of 0.20 V was produced in the low state. The amplifier was designed to drive the 120Ω load

of a braided ribbon cable at ECL levels (800 mVpp). However, to view the output an oscilloscope with a 50Ω termination was required. With no current flowing into the input of one of the amplifiers, the measured DC voltage on the input was 0.85 V. The threshold voltage for limiting amplification was crudely measured using a $1.0\text{ k}\Omega$ resistor connected to the input in order to simulate the impedance of a photodiode, it was found to be 0.025 V across the $1\text{ k}\Omega$ resistor. Ohm's law then dictates that the threshold current for these amplifiers is $25\text{ }\mu\text{A}$, close to the design value of $10\text{ }\mu\text{A}$. When the positive supply was increased to 3.2 V, the amplifier began to oscillate.

5.6 Integration of the Amplifier Chips into a Receiver Module

The amplifier chips were placed in a 196 pin ceramic flat-pack package with an externally supplied monolithic array of *p-i-n* photodiodes, as pictured in Figure 5-21. The photodiode array consisted of a rectangular grid of 8×8 photodiodes, and four amplifier chips were placed in the package with it, one adjacent to each side of the photodiode array. This hybrid packaging arrangement allowed issues such as amplifier crosstalk and prototype interconnections with an 8×8 VCSEL transmitter array to be investigated. The number of supplied amplifier chips was insufficient to allow testing of the amplifiers with the polysilicon MSM photodiode, and the priority was to test the amplifiers with known good, high-bandwidth photodiodes so that the performance of the amplifiers could be gauged. Moreover, the -3dB bandwidths of the polysilicon MSM photodiode and the

limiting amplifier were predicted to be so close that it would be difficult to measure the frequency response of each part with the two operating together in one module.

The analog frequency response of the array amplifier-photodiode module was measured using an Advantest R3361A Spectrum Analyzer (SA) and a laser diode transmitter. The same transmitter and a Broadband Communications Products model 110/210 Bit Error Rate tester (BERT) was used to test the digital performance. The laser diode in the transmitter was a Sharp model LT015MD, and was driven by either a frequency swept sinusoidal modulation from the SA or a pseudo-random bit sequence from the BERT. The package pins on the laser diode were cut to a length of approximately 1 mm and then soldered, together with a $47\ \Omega$ chip resistor, to an SMA connector which was subsequently connected to a bias tee (Picosecond Pulse Labs model 5555). The modulating signal was amplified by a Mini-Circuits ZPUL-21 $50\ \Omega$ amplifier and fed through the bias tee together with the appropriate DC bias to ensure full amplitude modulation of the transmitted optical signal. The laser output was coupled into approximately 2 m of $62.5\ \mu\text{m}$ core diameter multi-mode optical fibre and used to illuminate selected photodetectors on the 8×8 array or a reference photodiode.

Before measuring the analog frequency response of the packaged receiver, a high bandwidth reference photodiode [5.10] having a flat frequency response up to 6 GHz was first used to measure the frequency response of the ZPUL-21 amplifier/laser/fiber combination. The laser current was modulated with a sine wave, produced by the SA, that was swept from DC to as much as 1.3 GHz. The output from the reference photodiode was then fed directly back to the SA; Figure 5-22 shows the measured frequency response of the combination of the ZPUL-21 amplifier, laser, and length of

Figure 5-21: Photograph of the externally supplied photodetector array together with four 1x16 amplifier chips in a 128 I/O high-speed ceramic package. Each of the four chips is approximately $3.4 \times 1.5 \text{ mm}^2$. The output of each photodetector is wirebonded to an amplifier input.

optical fiber. The laser diode transmitter has a -3 dB bandwidth of $\sim 150 \text{ MHz}$ and is 10 dB down at 500 MHz . The ripples seen in the transmitter frequency response are due to interference caused by reflections at each end of the length of fiber used to carry the optical power. The resonance frequency spacing of approximately 30 MHz matches that produced by a Fabry-Perot etalon having the same length and refractive index as the fiber. The resonant peaks are not seen in Figures 5-23 and 5-25 because in those cases the signal is amplified by a non-linear amplifier (the limiting amplifier) and so the peaks are not amplified linearly.

Figure 5-22: Measured frequency response of laser diode transmitter measured with a known high bandwidth (>6GHz), flat frequency response, DC coupled photodetector.

The frequency response of the packaged receiver module, including 8x8 photodetector array and limiting amplifier was measured in the same manner as was the transmitter, using the transmitter as the stimulus. The photodetector-amplifier module was mounted on a circuit board with a short line to a 50 Ω cable; the cable was fed to another ZPUL-21 amplifier whose output was connected to the SA input. This second ZPUL-21 amplifier was used primarily to drive the signal into a 50 Ω environment. The 50 Ω cable from the receiver module to the ZPUL-21 was cut to a length of 2 cm to minimize the effects of impedance mismatch. Figure 5-23 gives the measured analog frequency response of the entire ZPUL-21/laser diode/fiber/photodetector-amplifier module/ZPUL-21 combination. When the strong frequency roll-off of the transmitter from Figure 5-22 is considered, the frequency response of the receiver module alone effectively falls slowly with increasing frequency and has a -3 dB bandwidth of approximately 650 MHz. The bandwidth of a receiver amplifier must be carefully chosen in order to maximize the performance of the receiver. The bandwidth must be sufficient

to pass the signal, but additional bandwidth will pass extra noise. For this reason, it is desirable to have an amplifier bandwidth that is just sufficient to pass the signal and no greater. The 650 MHz bandwidth of the measured receiver module is sufficient to pass 622 Mbit/s NRZ data with little excess bandwidth to pass additional noise. Figure 5-24 shows the traces of Figure 5-22 and Figure 5-23 together for comparison.

Crosstalk to an adjacent amplifier on the 1x16 amplifier chip was also investigated. The signal from a non-illuminated photodetector-connected amplifier adjacent to an active one is given in Figure 5-25. The frequency response presented in Figure 5-25 represents the crosstalk from a channel producing the output given in Figure 5-23 to an adjacent channel. The crosstalk level varies with frequency and typically ranges between -30 and -40 dB compared with that of the active channel.

Figure 5-23: Measured frequency response of the optical link containing the transmitter, limiting amplifier, and photodetector array. The response was obtained using the transmitter whose frequency response was given in Figure 5-22.

Figure 5-24: Measured frequency response of the (a) optical system containing the transmitter and receiver, and (b) laser diode transmitter measured with a known high bandwidth (>6GHz), flat frequency response, DC coupled photodetector. Subtracting the response in (b) from that in (a), gives a -3 dB bandwidth of approximately 650 MHz which is the frequency response of the receiver alone.

Figure 5-25: Measured frequency response for total signal crosstalk (optical and electrical) from the active channel to an amplifier immediately adjacent to the active amplifier, on the same 1x16 amplifier chip. The crosstalk varies in amplitude, but is typically between 30 and 40 dB down from the active channel signal.

The laser diode was driven with a pseudo random bit sequence of 127 bit word length at both 311 Mbit/s and 622 Mbit/s and used to characterize the digital performance of the amplifier/photodetector module. Figure 5-26 shows the measured eye at 311 Mbit/s

from the ZPUL-21/laser/fiber/photodetector-amplifier module/ZPUL-21 combination for one of the 64 photodetector/amplifier outputs. The voltage difference, shown by Figure 5-26, between the high and low states of the eye is approximately 11 divisions at 160 mV/division. Factoring in the 21 dB of gain provided by the ZPUL-21 gives a signal level of 160 mV. The amount of eye opening shown in Figure 5-26, defined as the difference between the minimum high state voltage and the maximum low state voltage, is slightly less at 120 mV. Figure 5-27 shows the output eye at a bit rate of 622 Mbit/s using the same arrangement. Again Figure 5-27 shows a signal level of 160 mV, but the amount of eye opening is somewhat less at 70 mV. One reason for the eye closure seen in Figure 5-27 is the limited frequency response of the laser diode transmitter used in the test arrangement (see Figure 5-22). The eye diagram of Figure 5-28 shows the optical output from the amplifier/laser/fiber received by the high bandwidth reference photodiode

Figure 5-26: Receiver module output eye at a bit rate of 311 Mbit/s (top) amplified by a Mini-Circuits ZPUL-21 amplifier, and 311 MHz system clock (bottom). The time scale is 1 ns per division and the voltage scale for the eye diagram is 160 mV per division, measured eye jitter is 400ps.

[5.10]. The rise time of the signal as shown in both Figures 5-26 and 5-27 is notably longer than the fall time for the same signal. This discrepancy was due mainly to the asymmetrical current drive capability of the output stage of the amplifier/photodetector module. The limiting amplifier circuit, shown in Figure 5-18, has an output stage that can sink more current through the output transistor than it can source through the diode connected npn transistor in series with the resistor. Another cause for the asymmetry in rise and fall times seen in the eye of Figure 5-27 is the slightly asymmetrical input eye produced by the ZPUL-21/laser diode combination as shown in Figure 5-28. Since the limiting amplifier inverts the input signal, the asymmetry shown in Figure 5-27 is the inverse of that shown in Figure 5-28.

The measured bit error rate plotted as a function of the optical power incident on one of the 8x8 array photodiodes is given in Figure 5-29. The system including transmitter and receiver module achieves a bit error rate of $<10^{-12}$ over a range of incident average power levels from 30 to 250 μW , this represents a dynamic range of 9.2 dB. Increasing the incident optical power, at low incident power levels, causes the signal to noise ratio to increase and results in a decreased bit error rate. At relatively high incident optical power levels, on the other hand, the input signal mid voltage level far exceeds the threshold level of the limiting amplifier. By increasing the incident optical power level, the threshold level essentially moves further away from the mid voltage level and toward the low voltage level of the input signal. Given that at maximum bit rate an input signal consisting of a 1010 pattern will approximate a sine wave, a slicing threshold that is not at the mid voltage level of the input signal will cause the duty cycle of the output signal to shift from the optimum value. A slicing threshold that is too high will cause the '1' bits

to be too narrow and the '0' bits to be too wide (in time), while a slicing threshold that is too low will cause the '1' bits to be too wide and the '0' bits to be too narrow. In this way as the incident optical power increases, the duty cycle of the output signal shifts so that the '1' bits grow ever wider at the expense of the '0' bits which become more narrow. While this is happening input signal noise is increasing while the amount of signal going into the '0' bits is essentially decreasing. Hence the input signal to noise ratio can actually fall with increasing incident optical power when a fixed slicing threshold receiver is employed. This highlights one of the limitations of limiting amplifiers, which generally have fixed slicing thresholds.

This receiver was designed to be effective for use in short optical interconnect distances, where optical attenuation is less than 10 dB, rather than to be highly sensitive. At 622 Mbit/s operation a measured minimum of 30 μW average incident optical power was needed to produce a bit error rate of $< 10^{-12}$, and this corresponds to an optical switching energy of 96 fJ.

Figure 5-27: Receiver module output eye at a bit rate of 622 Mbit/s (top) amplified by a Mini-Circuits ZPUL-21 amplifier, and 622 MHz system clock (bottom). The time scale is 500 ps per division and the voltage scale for the eye diagram is 160 mV per division, measured eye jitter is 400ps.

Figure 5-28: Transmitted eye as seen by high-speed reference photodiode. The time scale is 500 ps/div and the voltage scale for the eye is 15 mV/div. Measured jitter for the transmitted eye is 250 ps. Note the asymmetry in rise and fall times of the eye.

Figure 5-29: Bit error rate (BER) plotted as a function of average optical power incident on one of the 8x8 photodetectors at 622 Mbit/s. The two curves represent the BER performance of the receiver in sensitivity and overload.

CHAPTER 6

INTEGRATION OF THE POLYSILICON MSM PHOTODIODE INTO A TYPICAL Si VLSI PROCESS

6.1 Introduction

While the simplicity of the polysilicon MSM photodiode described in Chapter 4 makes this photodetector ideally suited to integration into a silicon VLSI CMOS process, the majority of existing CMOS processes do not exhibit all of the process steps required for the inclusion of the polysilicon MSM photodiode. In this chapter, a typical silicon CMOS process will first be described and then altered slightly to allow for the integration of the polysilicon MSM photodiode into the process. The purpose of this exercise is to illustrate the ease with which this photodetector can be added to a typical silicon process flow.

6.2 CMOS Process Sequences

Many variations of CMOS process flow exist; each one is different and no standard process has been, or likely ever will be, adapted by industry. Two CMOS process flows will be described in this chapter. The first describes a simple single-well,

single-level-metal CMOS flow, similar to the earliest CMOS processes adapted by industry. The second illustrates a modern twin-well, multi-level-metal process possessing characteristics such as lightly doped drains, trench isolation, vias and contact holes filled with metal, and dielectric planarization. The former process is typical of CMOS circuits having a minimum feature size of more than 1 micron, while the latter is exemplary of circuits with submicron dimensions.

6.2.1 Single p - and n -Well Process

A simple single-well CMOS process can be designed using either p -wells or n -wells and eight masking levels. The eight masking steps of a basic p -well process are illustrated in Figure 6-1. A typical n -well CMOS process begins with either a lightly doped $\langle 100 \rangle$ oriented ($\leq 10^{15} \text{ cm}^{-3}$) p -type wafer, or a heavily doped $\langle 100 \rangle p^+$ wafer having a thin (5-10 μm thick) lightly doped p -type epitaxial layer at the surface. The NMOS devices in an NMOS process are formed in the substrate, whereas the PMOS devices are placed within the more heavily doped ($\sim 10^{16} \text{ cm}^{-3}$) n -wells which are formed in the top surface of the substrate.

The n -wells are the first feature to be created on the starting wafer. The n -wells are formed by first growing a thermal oxide layer on the surface of the wafer, on top of which a CVD nitride film is deposited. The first mask, *mask #1*, is used to open windows in this layer, through which phosphorus is implanted.

The wafer is annealed in subsequent high-temperature cycles, allowing the implanted ions to diffuse to the appropriate depth, creating the n -well. Also during the drive-in stage, an oxide layer is grown over the n -well regions of the wafer. Following n -well creation, another ion-implant step is carried out whereby boron ions are implanted in

Figure 6-1: Cross-sectional views of a typical single-well CMOS process along with top views of each mask for every mask step involved in the process. The masks are in order: (a) p -well mask; (b) active-area mask; (c) polysilicon gate mask; (d) p^+ mask; (e) n^+ mask; (f) contact mask; (g) metal mask; (h) passivation mask.

the n -well areas. The purpose of this implant is to adjust the threshold voltage of both the NMOS and PMOS devices.

In the next step the previously introduced oxide and nitride layers are removed, and a fresh oxide/nitride layer is formed on the wafer surface. The wafer is then patterned using *mask #2* to define the active device areas and field regions. This photolithographic step leaves the oxide/nitride layers only over the regions which are to become drains, sources, channel regions, and diffusion interconnects. Without the use of a separate mask, a boron channel-stop implant is then carried out to enhance the properties of the p -substrate field regions. The concentration of boron used in the p -type channel-stop implant is not high enough to significantly alter the properties of the phosphorus-doped n -wells. The surface concentration of phosphorus in the well remains sufficiently high that a separate channel-stop is not necessary for the n -well regions.

Following the channel-stop implant, a thick field oxide layer is grown by localized oxidation of silicon (LOCOS). The previously deposited layer of silicon nitride acts as a mask against the thermal oxidation of the underlying silicon, and so by patterning the nitride layer, oxide can be selectively grown to match the pattern. Next the nitride/oxide layers which were used to define the active device regions are removed, and a gate oxide layer is grown. A polysilicon gate layer is then deposited by chemical vapour deposition (CVD) to a typical thickness of 0.4-0.5 μm [6.3], and subsequently doped to solid solubility with phosphorus to produce an n -type polysilicon gate. The third mask, *mask #3*, is used to pattern the polysilicon layer. Since this layer serves as an interconnect as well as a gate, the polysilicon must be fairly heavily doped.

The source/drain regions of the active devices are the next features to be formed. *Mask #4* is used to implant the source/drain of the PMOS devices, while *mask #5* is employed to implant the source/drain of the NMOS devices. The device channel region, which lies under the gate, is shielded from the source/drain implant by the overlying polysilicon gate. Both the source and drain regions need to exhibit low resistivities; typical values of source/drain sheet resistance should be less than $30 \Omega/\square$. In order to obtain sufficiently low resistivities, the dopant concentration must be sufficiently high. A combination of arsenic and phosphorous is typically used to dope the n^+ regions, and it is typically implanted to a dose of $3\text{-}6 \times 10^{15} \text{ cm}^{-3}$. The n^+ dopant is implanted with a low energy (40-60 keV) so that it stays near the surface. For the p^+ regions, BF_2^+ is often used as the dopant to produce shallow junctions which do not diffuse much. It is typically implanted to a dose of $1\text{-}5 \times 10^{15} \text{ cm}^{-3}$ with energies of 30-50 keV. The dopant concentration in the source/drain regions of both the NMOS and PMOS devices is sufficiently high that when metal is later deposited on these source/drain regions, ohmic contacts are formed. Typically, a concentration of $>10^{20} \text{ cm}^{-3}$ is used to dope these areas. After the source/drain implants are performed, the wafer is annealed to a moderately high temperature (900-1000° C) to diffuse the dopants. Following the anneal step, it is typical for the gate oxide to be removed and then regrown, as up to this point it has suffered damage caused by the heavy ion implants performed.

The next step in the process is to deposit a CVD oxide layer on the wafer surface to a depth of 50-100 nm. This oxide layer is subsequently doped with boron and/or phosphorus, and then flowed [6.1] to smooth out the wafer surface and allow for better metal step coverage. The purpose of this oxide layer is to serve as both a dielectric

between polysilicon and metal layers, as well as a gettering layer to reduce the level of metallic impurities and crystalline defects in the silicon substrate [6.2].

The next mask, *mask #6*, is used to open contact windows in the layer of CVD oxide. These windows in the oxide are formed so that metal is able to contact the source and drain regions of the active devices. Next, an Al or an Al-Si alloy is deposited and then patterned using *mask #7*. The wafer is then annealed in a hydrogen atmosphere to passivate defect sites, and then the final oxide layer (often called a passivation layer) is deposited. *Mask #8* is used to open windows in the passivation layer to form bond pads.

6.2.2 Advanced Twin-Well CMOS Process

The process sequence described in this section is a more advanced, twin-well silicon process. This process closely follows Intel's 0.25 μm Si process [6.4] and is intended to represent a modern twin-well CMOS process. This process sequence builds on the previously described single-well CMOS process and as such, details of process steps which are similar to those of the single-well process will be omitted. Only those steps that are not present in the single-well process will be described in detail.

A typical twin-well CMOS process begins with a $\langle 100 \rangle$ oriented, heavily doped wafer on top of which a thin (5-10 μm thick) lightly doped epitaxial layer has been grown. The most common configuration is a p^+ substrate under a p -type epitaxial layer. The n -epitaxial layer on n -substrate is however used for a few special applications, but this configuration is more sensitive to process-induced defects [6.3].

The small geometries involved in an advanced twin well silicon process leads to devices that are so closely spaced that electrical isolation becomes an issue. It can be very beneficial to process performance to isolate most components, interconnecting them only

with the metallization pattern. The process described here includes a step for creating shallow trench isolation; these features are formed before the twin wells. If trenches are not used the process is less complex and requires one less masking step. The process of making shallow isolation trenches begins with the deposition of a nitride layer over a bare silicon wafer, this layer is then patterned and used for an etch mask for an anisotropic etch into the silicon to form the trench. A deep, narrow trench having nearly vertical sidewalls can be formed using a reactive ion etch machine. The trench surface is oxidized and then filled in with polysilicon by chemical vapour deposition. The wafer surface can then be etched flat and oxidized, producing a planar surface over the newly formed trench. Electrical isolation occurs when the trench is deep enough to penetrate through the epitaxial layer to reach the underlying substrate. In this case the oxide walls prevent carriers from penetrating through the trench, and the back to back diodes formed by the epitaxial layer to substrate contact in two different isolated regions prevent carriers from going under the trench.

The creation of twin wells is the next step in the process. The most widely used method of creating the twin wells relies on only a single mask. This method uses one mask to pattern a nitride/oxide layer after it has been deposited on the wafer surface. Phosphorus is implanted into the n -well regions through windows in the nitride/oxide layer, following which a thermal oxide is grown on top of these regions. The oxide must be thick enough to block the subsequent boron implant that is performed following the removal of the nitride/oxide layer. This boron implant creates p -well regions everywhere on the wafer surface that an n -well does not already exist.

As in the single-well CMOS process, following well implantation the dopants must be driven into the substrate by exposure to high temperatures for lengthy periods. Dopant concentration in the n -well regions can be made higher than it is in p -well regions ($\sim 3 \times 10^{16} \text{ cm}^{-3}$ vs. $\sim 1 \times 10^{16} \text{ cm}^{-3}$) to eliminate the need for a separate channel-stop implant for the n -well regions, and to improve the punchthrough performance of the PMOS devices. If this is the case, the boron channel-stop implant is performed following the well implant steps, without the use of an additional mask.

When the dopant concentration in the n -well regions is not sufficiently high to omit the n -well channel stop, an additional mask is used to implant both n -well and p -well channels. Here, the boron channel stop implant is introduced into both the n -well and p -well field regions. The phosphorus channel stop however, is implanted only into the n -well field regions, and must be increased to compensate for the presence of boron. Drawbacks of this method range from the additional requirement of aligning the extra mask, to the interdiffusion of dopants that occurs during subsequent exposures to high-temperatures. A maskless variation of this process, however, has been developed to overcome these disadvantages. This variation uses both arsenic and phosphorus to dope the n -well regions (as defined by *mask #1*); one is used to form the well while the other is used to form the channel stop. An oxide is then grown on top of the n -well regions to mask a subsequent boron implant that is carried out for the p -well regions. After the wells have been driven in, the oxide over the n -well regions remains while a second boron implant step is performed. This second boron implant acts as a channel stop in the p -well field regions as well as a punchthrough prevention implant in the active areas of the p -well regions.

Field oxide growth is normally performed following the channel stop implant steps. The high temperatures associated with field oxide growth, however, tend to cause diffusion of the boron channel stop implant. One method developed to combat the excessive redistribution of boron during field oxide growth relies on an elevated ambient pressure to reduce the anneal time required to grow the field oxide. This method has the added benefit of reducing the amount of boron channel stop dopant that is lost to the growing field oxide layer. Moreover, this method causes less well dopant diffusion, allowing n - and p -channel devices to be placed in closer proximity.

The next steps in the modern twin-well CMOS process are to define and separate the active and field regions, followed by gate definition. Localized oxidation of silicon (LOCOS) is used to grow a field oxide that reduces device size and improves device performance. Following active region definition, a gate oxide layer is grown on the wafer. Intel uses a gate oxide thickness of 4.08 nm [6.4] in their 0.25 μm silicon process. The next step is to perform a threshold-voltage adjust implant. This can be accomplished with either a single implant procedure, or with two separate implant steps requiring an additional mask; the single step implant is performed with boron. For submicron devices, a separate selective punchthrough-prevention implant for NMOS devices is often required. This step also requires an additional mask. These threshold-voltage adjust and punchthrough-prevention implants can also be performed before the gate oxide is grown. But in order to keep the boron threshold-voltage implant shallow and improve the PMOS punchthrough characteristics, the dopants are often implanted through the gate oxide layer. A partial polysilicon deposition is often performed before the implant step in order to

protect the oxide layer from contamination, the remainder of the polysilicon layer is then deposited following completion of the threshold-voltage implant.

Polysilicon deposition is the next procedure in the process, after which the polysilicon is patterned to form device gates. Generally, the procedure for deposition and patterning of the polysilicon is similar that used in the single-well CMOS process. For a modern twin-well process however, a *salicide* process is often used on the polysilicon, source, and drain regions in order to reduce the parasitic resistance of these regions. A salicide process involves the deposition of a refractory metal, masked by the field oxide, following the source/drain implant steps. The wafer is then heated so that a silicide layer is formed on the surface of the polysilicon and source/drain. Because the silicide is formed wherever the deposited metal touches exposed silicon, the process has been coined self-aligned silicide, or salicide. Intel's 0.25 μm process uses titanium to form Ti Si_2 gate, source, and drain regions [6.4].

Once the polysilicon has been deposited, it is then patterned. Because gate lengths can be very small (in the submicron range), the etching process must be very accurate. The etch must also produce vertical sidewalls in order to achieve greater performance in devices possessing lightly doped drain structures (see below), as such the etch must be highly anisotropic. Moreover, the etch must be very ineffective against removing the thin layer of gate oxide which lies beneath the polysilicon layer.

The source and drain junctions required for submicron CMOS devices need to be substantially shallower than those used in larger devices in order to minimize short-channel effects. Following polysilicon gate definition, Intel's 0.25 μm process includes a silicon implant to create an amorphous layer in the gate, source, and drain regions of both

n-type and *p*-type devices. The purpose of the amorphous layer is to limit the spread of dopants from subsequent implant steps, and results in more abrupt dopant profiles. The elimination of the crystal lattice in the substrate reduces channeling and helps to form more shallow junctions when the dopants are implanted. By reducing the lateral implant tails under polysilicon gate regions, sub-threshold leakage in short channel devices can be minimized [6.4]. The energy and dose of the silicon implant must be sufficiently high to amorphize the polysilicon gate (as well as the source/drain regions) without substantially damaging the underlying gate oxide [6.4].

The formation of the source and drain regions in advanced CMOS processes must differ somewhat from that in more basic CMOS processes. Since the gate lengths tend to be so small in advanced CMOS processes short-channel effects can become noticeable, especially in NMOS devices, as device geometries shrink faster than device voltages. As channel lengths shrink in size, the depletion regions of the source and drain move closer together and eventually begin to encroach into the channel area, even without an applied bias. The result is *charge sharing* by the source and drain of the charge intended for the channel, note that this charge is no longer under the control of the gate voltage. When a bias is applied to the drain, the drain depletion region extends even further into the channel area and can reduce the effective doping level of the channel area near the source. When this happens, the potential barrier between the source and drain is lowered and carriers are easily injected into the channel. At this point the gate voltage no longer has control of the drain current.

Another result of smaller device geometries is the increase in electron energy levels known as the hot electron effect. This effect is a result of the rise in electric field

strength caused by the reduced distances between conductors. The electric field in the reverse-biased drain junction of an n -channel device can become high enough in small geometry devices that impact ionization occurs. The resulting hole current contributes to the substrate current, and some of this current may reach the source junction where they contribute to the lowering of the source barrier. The effect of a reduced source barrier is to allow for more electrons injected from the source into the p -region. This n - p - n transistor action within the source-channel-drain can become so bad that control of the device current by the gate is no longer possible. Another hot electron effect occurs when electrons gain sufficient energy to be carried over or tunnel through the barrier formed by the gate oxide and become trapped within the oxide layer. Here they change the device threshold voltage and I - V characteristics.

The reduction of short channel and hot carrier effects can be accomplished by lowering the doping concentrations of the source and drain regions so that the junction fields are lower. However, source/drain regions that have low dopant concentrations tend to also have high series resistance. A solution is to produce shallow, lightly doped source/drain extensions nearest to the channel, while doping the remaining source/drain regions as before. This lightly doped drain (LDD) structure produces lower electric field between the drain and channel regions, thereby reducing hot carrier effects like carrier injection into the oxide and impact ionization.

The formation of LDD devices relies on the use of *sidewall spacers* on either side of the transistor gate. Following the formation of the polysilicon gate a low-dose n -type implant is used to form shallow, lightly doped source and drain extensions in an NMOS device. Next a thick layer of nitride or oxide is deposited and is then etched away by

reactive ion etching. The remaining dielectric from the previous deposition has rounded ends that butt up against the vertical walls of the polysilicon gate; these are the sidewall spacers. A second source/drain implant is then carried out whereby high concentrations of dopants are implanted, masked by the gate and sidewall spacers, and then driven in with a diffusion step. The resulting profile consists of heavily doped source/drain regions separated from the channel by small lightly doped source/drain extensions. A similar set of steps is required to produce LDD structures in PMOS devices. Processes having gate widths of less than $0.5\text{ }\mu\text{m}$ typically use the source/drain implant step to simultaneously dope the polysilicon gate. The polysilicon gate material used for NMOS devices is doped to form *n*-type polysilicon, while *p*-type polysilicon is used for PMOS devices. Intel's $0.25\text{ }\mu\text{m}$ process uses complimentary doped polysilicon gate material to obtain matched threshold voltages in NMOS and PMOS devices [6.4].

A halo implant is used in Intel's $0.25\text{ }\mu\text{m}$ process and others in the same lithography step used to dope the source/drain extensions. The halo implant is a high-angle implant that must be performed in four 90-degree rotations to ensure that both sides of the transistor receive the implant as well as transistors oriented in both mutually orthogonal directions [6.4]. The same dopants that are used in the original wells are used for the halo implants. The device threshold voltage is determined by the halo implant together with the original well implant. By using less initial well dopant and performing the halo implant after gate definition, channel regions near source/drain extensions receive a higher dopant concentration than regions closer to the center of the channel. In this way, short channel devices receive a higher dopant concentration than longer channel devices. The benefits of the halo implant include a reduced threshold voltage roll-off in short

channel devices, and a higher I_{dsat} since a halo device has a more abrupt drain-channel junction and a higher channel mobility than a device without the halo implant [6.4].

To reduce junction capacitance, compensation implants can be used, typically performed during the same lithography step used for the source/drain implants. The compensation implant uses the same type of dopants used for the source/drain implant but with a lower dose and higher energy so that a more graded implant profile is formed at the junction with the underlying well. A compensation implant is capable of reducing junction capacitance by 20-30% with no degradation of device isolation or implant penetration of the gate oxide [6.4]. This reduction of junction capacitance greatly helps the high-frequency performance of the process.

Following the formation of the source and drain regions, an oxide layer is deposited to isolate the substrate from the overlying metal layer to be applied in an upcoming step. The oxide used under metal 1 in the Intel process is Boro-Phosphosilicate-Glass (BPSG), and is planarized using chemical-mechanical polishing (CMP). A contact mask is used to open windows in this oxide layer to allow for electrical contact from the source, drain, gate, and substrate regions to metal 1, the bottom-most metal layer. Contact holes and vias are filled with tungsten plugs, formed by blanket deposition of tungsten by CVD followed by CMP to remove all tungsten that overlies BPSG. Tungsten is used because of its excellent step coverage characteristics, although polysilicon can also be used.

The first metal layer is applied following contact formation. Intel's 0.25 μm process uses five metal layers that vary in thickness but are each comprised of a stack of metals consisting of Ti/Al-Cu/Ti/TiN in order from lowest to highest [6.4]. After each

metal stack is deposited and patterned, another dielectric layer is deposited, etched back, and then planarized using CMP. It is important to planarize this layer to avoid gaps or breaks in the metallization, which can occur where the oxide layer is particularly steep. Several methods have been used to achieve planarization, but they all succeed in replacing any steep slopes in the dielectric surface with more gentle slopes. After planarization, windows in the dielectric must be opened in order to establish contact between the upper and lower metals. A mask is used to define these contact points called vias which, once formed are then filled in with tungsten plugs. This process is repeated until the final metal has been deposited and patterned. The final step in the twin-well process is the passivation of the surface by performing a hydrogen anneal as in the basic CMOS process, followed by deposition and patterning of the passivation layer, which also proceeds in the same manner as described in the basic CMOS process.

6.3 Incorporating Polysilicon MSM Detectors into a Si Process

While the 0.25 μm twin-well silicon processes described in the previous section is not intended for the formation of polysilicon MSM photodiodes, small changes to these processes will allow their incorporation with little added process complexity. Regardless of the fabrication process used, a layer of undoped polysilicon must be deposited on top of a layer of material that does not produce and conduct photogenerated carriers, a dielectric layer will work. Another requirement is for an optically low-loss, line of sight path to the top surface of the wafer on which the photodiode is fabricated. This means that no metal layers can be placed directly over the fabricated photodiode, and that any

dielectric placed over the photodiode should be thin and highly transmissive to optical wavelengths that lie within the sensitivity band of the photodiode.

Many of the building blocks of the polysilicon MSM photodiode (polysilicon, metallization, a Schottky barrier, and a dielectric layer) are already used in the vast majority of silicon CMOS processes. All that is needed to produce a polysilicon MSM photodiode in a standard silicon process is to properly arrange these materials. The addition of a polysilicon MSM photodiode to Intel's 0.25 μm twin-well process [6.4] begins with the deposition and patterning of an undoped, or lightly doped, polysilicon layer on top of oxide. The polysilicon could come from the gate polysilicon, or a separate polysilicon deposition could be performed. In either case an additional mask is required to form the MSM photodiode. When using gate polysilicon a mask is necessary to protect the photodiode regions from the dopants applied to the gate polysilicon and from the salicide process. If, on the other hand, a separate polysilicon deposition is used to form the base of the MSM photodiode, a mask is needed to pattern the polysilicon, Figure 6-2 illustrates the additional steps. Contact windows are then opened over the polysilicon regions intended for MSM photodiode use. The same contact mask that is used in Intel's 0.25 μm twin-well process can be used for the MSM photodiode because the metallization stack Intel uses makes good Schottky contacts. The base layer in the metallization stack used by Intel is titanium [6.5], which matches the base of the metallization stack used to fabricate the polysilicon MSM photodiode as described in Chapter 4.

Once the metallization step has been performed, all that is left to do is ensure that most, if not all, of the dielectric formed above the photodiode has been removed by

localized etching. A separate mask will be needed to perform this etch step. It may prove beneficial to ensure that the photodiode surface is passivated by a hydrogen anneal along with the rest of the silicon wafer by performing this step after the removal of dielectric above the photodiode. The extent to which the preceding statement is true however, has yet to be determined and requires further research. It must be stated for clarity that the addition of a mask to a silicon IC fabrication process is costly. The owner of a process will only undertake the costly inclusion of additional steps to the process if an adequate return on investment is predicted. The return on investment in the addition of extra mask steps to a silicon process will not be adequate unless sufficient demand exists for the features that the extra steps bring to the process. While the demand for integrated optical detectors is rising, it will most likely be several years before sufficient demand exists to justify the addition of an optical detector to a commercial silicon IC fabrication process.

(a)

(b)

Figure 6-2: (a) Schematic cross section of Intel's 0.25 μm CMOS process up to the point just prior to deposition of first level metal. (b) The same wafer after an additional mask step that would add the polysilicon MSM photodiode to the process, showing patterned polysilicon with first level metal and planarized oxide over top.

CHAPTER 7

CONCLUSION

7.1 Thesis Conclusions

This thesis describes the research, design, fabrication, and test of two components of an optical communications link. The components that were produced include a silicon MSM photodiode that is VLSI compatible, and a microelectronic limiting amplifier, fabricated in a silicon VLSI process, designed to amplify the small electrical signal produced by a photodiode.

7.1.1 Silicon-based Photodiode

Optical interconnections are a cost-effective solution for high bandwidth interconnections over optical fibre for computer networking. One of the required elements in an optical backplane or interconnect is a low-cost receiver, and the integration of Si-based detectors with Si VLSI circuitry is an active area of current research and development [1.23 – 126, 2.5 – 2.12].

While silicon photodetectors can be relatively inexpensive to produce, their performance tends to be low. This is because the large absorption depth of silicon causes

the creation of charge carriers in regions far from the device electrodes where the electric fields are weak. Here, the photogenerated carriers either slowly drift or diffuse toward the electrodes, and produce a signal with a very long rise and fall time.

In this thesis a novel Si-based MSM detector is described [1.28] which offers good performance characteristics and was fabricated using simple, low-cost, Si VLSI compatible fabrication processes. To the author's knowledge this detector represents the first use of polysilicon as the photoactive material in a metal-semiconductor-metal (MSM) photodiode. As an MSM structure it is a lateral device, and has the advantage of simpler fabrication than p-i-n devices, which are typically vertical structures. Furthermore, polysilicon absorbs strongly in the 800-860 nm range, which is compatible with readily available GaAs laser sources. The detector makes use of a thin oxide layer to electrically isolate the polysilicon layer from the silicon substrate, thereby ensuring that all carriers are generated within a high-field region.

Table 7-1 gives the measured performance metrics for the polysilicon MSM photodiode. The data in Table 7-1 can be easily compared with that of Table 2-1, so that the performance of the polysilicon MSM photodiode can be compared with the other recently published photodiodes for use with 850 nm optical signals that are summarized within Table 2-1. The polysilicon MSM photodiode offers performance characteristics that are on par with the photodiodes summarized in Table 2-1, yet should be easier to fabricate than most of them due to its simple construction.

The absorption coefficient of polysilicon is higher than that of monocrystalline silicon, and a 2 μm thick layer is sufficient to absorb ~60% of the incident light, giving the detector a measured responsivity of 0.13 A/W for 860 nm wavelength incident light.

The 2 μm minimum feature size simple silicon process that was used produced a device having a -3 dB bandwidth of 750 MHz. In contrast to similar detectors fabricated on bulk silicon, the device exhibits no extended current tail in its impulse response, giving it a significantly higher bandwidth. Improvement in device bandwidth can be accomplished by reducing either the contact spacing or the polysilicon layer thickness, the latter coming at the expense of device responsivity. Device responsivity can be improved by reducing reflections off the top surface, or by increasing reflections from the back surface. Ways of reducing reflections include reducing the contact finger width, and adding an antireflection coating. Increased reflection off the back surface can be accomplished by adding a reflecting structure such as a metallic layer, or a quarter wavelength stack (also known as a Bragg mirror) similar to those used in the fabrication of vertical cavity surface emitting lasers.

Table 7-1: Performance metrics of the polysilicon MSM photodiode

Technology	P.D. type	Dark Current (nA)	P. D. Area (μm^2)	Cap. (fF)	Responsivity (A/W)	-3dB Bandwidth (GHz)	Ref.
Silicon	MSM	1 000	10^4	100	0.13	0.75	[1.28]

Polysilicon is an integral part of standard Si VLSI structures where it is used to form resistors, and gate electrodes for MOS devices, and capacitors, and as such is already available for use in nearly all existing Si-VLSI processes. The good performance characteristics of this device together with its compatibility with existing VLSI processes

make the polysilicon MSM photodetector suitable for use in high-speed optoelectronic integrated circuits fabricated in silicon VLSI technologies.

7.1.2 Amplifier for Optical Receiver

Receivers for optical communication links require an amplifier to transform the weak signal produced by a photodiode to one whose amplitude is compatible with the circuitry which follows the amplifier. Moreover, the receiver must include a preamplifier that amplifies the weak signal produced by the system photodetector while adding as little noise to the signal as possible. An optical link which contains many optical channels must have a low per channel power consumption to keep the power consumption of the whole link at a reasonable level. It is therefore important to use in this type of link a receiver amplifier which has sufficiently low power consumption, and at the same time is fast enough to be compatible with the wide bandwidth signals it is designed to amplify.

The goal of this amplifier design was to achieve adequate performance for 622 Mbit/s operation at the lowest achievable power consumption. The power consumption of each amplifier together with its output driver stage was approximately 35 mW, and had an input dynamic range of greater than 9 dB over which a bit error rate of $< 10^{-12}$ was achieved. These performance characteristics compare well with other amplifier designs reported in the literature.

In order to save power, this receiver amplifier contains a main amplifier integrated with a preamplifier on a single die, contrary to the vast majority of amplifier designs which separate the main amplifier and preamplifier onto different dice. Integration of the two amplifiers saves power by eliminating the need of the preamplifier

to drive a $50\ \Omega$ load. The separation of the main and preamplifiers is generally done to minimize the chance of feedback from the output of the main amplifier to the input of the preamplifier. Feedback in the amplifier design presented here is reduced through careful design. Each stage of the amplifier has dedicated power supply lines that are connected off chip; the amplifier stages are arranged one after another in order to maintain a straight line signal path and to maximize the distance between the preamplifier stage and the final gain stage of the amplifier; substrate connected guard rings are placed around each amplifier stage in a star connection. All of the aforementioned practices act to reduce feedback within the amplifier and allow the integration of preamplifier and main amplifier.

Major design choices for this amplifier include the use of a limiting amplifier design over that of an automatic gain control design, and single-ended rather than differential gain stages. Limiting amplifiers are fixed-gain non-linear designs with sufficient gain to fully amplify the lowest expected input signal. These designs are simpler, and consume less power and chip area than automatic gain control amplifiers, which offer larger bandwidths, greater dynamic range, and lower bit error rate operation. The strict requirement on power consumption necessitated the use of single-ended gain stages which inherently use less power than differential gain stages. The advantages of differential gain stages include greater bandwidth, lower jitter, and greater noise immunity than single ended stages. These advantages come, however, at the price of larger power rail voltages, current and power consumption than those found in single ended circuits. In short the amplifier presented here has been designed with minimum power consumption and 622 Mb/s operation as the primary design goals. If greater noise

immunity and signal isolation, and lower signal jitter are required of the receiver however, it would be prudent to opt for an amplifier which uses differential circuitry.

Another unique feature of this amplifier is the use of a series-feedback stage as the output driver of the amplifier. The series-feedback output stage creates an impedance mismatch that allows the amplifier to drive low impedance loads at high frequencies thereby aiding in the creation of a high bandwidth circuit. By wire-bonding four of 1x16 amplifier chips to an 8x8 *p-i-n* photodiode array, a 64 channel receiver module was produced. The channels were measured to have a –3 dB bandwidth of 650 MHz, giving the module the capability to throughput 40 Gbit/s of NRZ data. Measured crosstalk to adjacent channels in the module was less than 30 dB, which is adequate for most optical interconnection systems. The next step in this research should be to integrate the polysilicon MSM photodiode onto a silicon VLSI process, then build and test a truly integrated optical receiver.

REFERENCES

- [1.1] "International technology roadmap for semiconductors," Semiconductor Industry Association, 2000.
- [1.2] Nan M. Jokerst, Martin A. Brooke, J. Laskar, D. Scott Wills, A. S. Brown, M. Vrazel, S. Jung, Y. Joo, and J. J. Chang, "Microsystem optoelectronic integration for mixed multisignal systems," *IEEE J. Select. Topics Quantum Electron.*, **6**, no. 6, pp.1231-1239, 2000.
- [1.3] J. D. Meindl, "Gigascale integration: is the sky the limit," *Circuits and Devices*, pp. 19-32, Nov. 1996.
- [1.4] D. A. B. Miller and H. M. Ozaktas, "Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture," *J. Parallel Distrib. Comput.*, **41**, p. 4252, 1997.
- [1.5] A. V. Krishnamoorthy and D. A. B. Miller, "Scaling optoelectronic-VLSI circuits into the 21st century: a technology roadmap," *IEEE J. Select. Topics Quantum Electron.*, **2**, no. 1, pp. 55-76, 1996.
- [1.6] D. A. B. Miller, "Rationale and challenges for optical interconnects to electrical chips," *Proc. IEEE*, **88**, no. 6, pp. 728-749, 2000.
- [1.7] R.A. Nordin et al., "A system perspective on digital interconnection technology," *IEEE J. Lightwave Technology*, **10**, pp. 811-827, 1992.
- [1.8] F. Kiamilev et al., "Performance comparison between optoelectronic and VLSI multistage interconnection networks," *IEEE J. Lightwave Technology* **9**, pp. 1665-1674, 1991.

- [1.9] H.S. Stone and J. Cocke, "Computer Architecture in the 1990s," *Computer* **24**, No. 9, pp. 30-38, 1991.
- [1.10] Ahmed Louri and Hongki Sung, "3D optical interconnects for high-speed interchip and interboard communications," *Computer* **27**, No. 10, pp. 27-37, 1994.
- [1.11] D.A.B. Miller, "Optics for low-energy communication inside digital processors: quantum detectors, sources, and modulators as efficient impedance converters," *Optics Lett.* **14**, pp. 146-148, 1989.
- [1.12] J. W. Goodman, "Optical interconnections in computing," in *Computation and Cognition. Proceedings of the First NEC Research Symposium*, pp. 22-31, 1991.
- [1.13] R. L. Khalil, L. R. McAdams, and J. W. Goodman, "Optical clock distribution for high speed computers," *SPIE* **991**, pp. 32-41, 1988.
- [1.14] Amal K. Ghosh, Charles Fishman, and Tom Feng, "Theory of the electrical and photovoltaic properties of polycrystalline silicon," *J. Applied Physics* **51** (1), pp. 446-454, 1980.
- [1.15] H. C. Lee and B. V. Zeghbroeck, "A novel high-speed silicon MSM photodetector operating at 830 nm wavelength," *IEEE Electron Device Lett.* **16**, pp. 175-177, 1995.
- [1.16] N. Streibl, K-H. Brenner, A. Huang, J. Jahns, J. Jewell, A.W. Lohmann, D.A.B. Miller, M. Murdocca, M.E. Prise, T. Sizer, "Digital optics," *Proceedings of the IEEE* **77**, pp. 1954-1969, 1989.
- [1.17] D.Z. Tsang, "One-gigabit per second free-space optical interconnection," *Appl. Optics* **29**, pp. 2034-2037, 1990.
- [1.18] E.E.E. Frietman, L. Dekker, S.A. Boothroyd, P. Palacharla, J. Chrostowski,

"Optics for multiple-broadcast massively-parallel and independently addressable free-space interconnections", in: Proc. 3rd IEEE International Workshop on Photonic Networks, Components and Applications, Atlanta 1993, pp.75-83 (ISBN 1-895634-04-0 OCRI Publications).

- [1.19] K.-H. Brenner, and F. Sauer, "Diffractive-reflective optical interconnects," *Appl. Optics* **27**, pp. 4251-4254, 1988.
- [1.20] R.P. MacDonald, S.A. Boothroyd, J. Chrostowski, B.A. Syrett, "Free-space intermodule interconnection using non-diffracting diode laser beams", in: Proc. Canadian Conf. on Electrical and Computer Engineering, Vancouver 1993, pp.1262-1265, (ISBN 0-7803-1443-3).
- [1.21] R.P. MacDonald, S.A. Boothroyd, J. Chrostowski, B.A. Syrett, T. Okamoto, "Optical interconnection schemes using diode laser Bessel beams", in: Proc. Canadian Conf. on Electrical and Computer Engineering, Halifax 1994, (Canadian Society for Electrical and Computer Engineering).
- [1.22] K. Hamanaka, "Optical bus interconnection system using selfoc lenses," *Optics Lett.* **16**, pp. 1222-1224, 1991.
- [1.23] K. Ayadi, M. Kuijk, P. Heremans, G. Bickel, G. Borghs, and R. Vounckx, "A monolithic optoelectronic receiver in standard 0.7- μm CMOS operating at 180 MHz and 176-fJ light input energy," *IEEE Photonics Tech. Lett.* **9**, pp. 88-90, 1997.
- [1.24] J. S. Wang, C. G. Shih, W. H. Chang, J. R. Middleton, P. J. Apostolakis, and M. Feng, "11 GHz bandwidth optical integrated receivers using GaAs MESFET and MSM technology," *IEEE Photonics Tech. Lett.* **5**, pp. 316-318, 1993.

- [1.25] J.-S. Reih, D. Klotzkin, O. Qasaimeh, L.-H. Lu, K. Yang, L. P. B. Katehi, P. Bhattacharya, and E. T. Croke, "Monolithically integrated SiGe-Si PIN-HBT front-end photoreceivers," *IEEE Photonics Tech. Lett.* **10**, pp.415-417, 1998.
- [1.26] Dennis L. Rogers, "Monolithic integration of a 3-Ghz detector/preamplifier using a refractory-gate, ion-implanted MESFET process," *IEEE Electron Dev. Lett.* **7** (11), pp. 600-602, 1986.
- [1.27] T. Y. Hsiang, S. Alexandrou, and C. C. Wang, "Picosecond silicon metal-semiconductor-metal photodiode," *Proc. SPIE* **2022**, 1993, pp. 76-82.
- [1.28] R. P. MacDonald, N. G. Tarr, B. A. Syrett, S. A. Boothroyd, and J. Chrostowski, "MSM photodetector fabricated on polycrystalline silicon," *IEEE Photonics Tech. Lett.* **11** (1), pp. 108-110, 1999.
- [1.29] R. P. MacDonald, S. A. Boothroyd, P. Myslinski, R. Misner, B. A. Syrett, and T. Kurokawa, "1 x 16 amplifier array and hybrid integration with an 8 x 8 *p-i-n* photodiode array for parallel optical interconnects," *Microwave and Optical Tech. Lett.*, **28** (1), pp. 11-15, 2001.
- [1.30] E.M. Cherry, D.E. Hooper, "The design of wide-band transistor feedback amplifiers," *Proc. IEE* **110** (2), pp. 375-389, 1963.
- [1.31] R. P. MacDonald, S. A. Boothroyd, T. Okamoto, J. Chrostowski, and B.A. Syrett, "Interboard optical data distribution by Bessel beam shadowing," *Optics Comm.* **122**, pp. 169-177, 1996.
- [2.1] Charles Kittel, Introduction to Solid State Physics, 6th Ed., John Wiley and Sons, Toronto, 1986.

- [2.2] S.M. Sze, Physics of Semiconductor Devices, 2nd Ed., John Wiley and Sons, Toronto, 1981.
- [2.3] Paul A. Tipler, Modern Physics, Worth Publishers, Inc. New York, 1978.
- [2.4] B.E.A. Saleh and M.C. Teich, Fundamentals of Photonics, John Wiley and Sons, Toronto, 1991.
- [2.5] Li-Hong Lai, Tien-Chang Chang, Yen-Ann Chen, Wen-Chin Tsay, and Jyh-Wong Hong, "Characteristics of MSM photodetectors with trench electrodes on p-type Si wafer," *IEEE Trans. Electron Dev.*, **45**, no. 9, pp. 2018-2023, 1998.
- [2.6] H. Zimmermann and T. Heide, "A monolithically integrated 1-Gb/s optical receiver in 1- μ m CMOS technology," *IEEE Photonics Tech. Lett.*, **13**, no. 7, pp. 711-713, 2001.
- [2.7] T. K. Woodward and A. V. Krishnamoorthy, "1 Gbit/s CMOS photoreceiver with integrated detector operating at 850 nm," *Electronics Lett.*, **34**, no. 12, pp. 1252-1253, 1998.
- [2.8] J. D. Schaub, R. Li, S. M. Csutak, and J. C. Campbell, "High-speed monolithic silicon photoreceivers on high resistivity and SOI substrates," *J. Lightwave Tech.*, **19**, no. 2, pp. 272-278, 2001.
- [2.9] S. J. Koester, B.-U. Klepser, J. O. Chu, D. Kuchta, and K. Ismail, "1.1 GHz MSM photodiodes on relaxed $\text{Si}_{1-x}\text{Ge}_x$ grown by UHV-CVD," *Proc. 56th Annual Device Research Conference*, pp. 60-61, 1998.
- [2.10] J.-S. Rieh, D. Klotzkin, O. Qasaimeh, L.-H. Lu, K. Yang, L. P. B. Katehi, P. Bhattacharya, and E. T. Croke, "Monolithically integrated SiGe-Si PIN-HBT

- front-end photoreceivers," *IEEE Photonics Tech. Lett.*, **10**, no. 3, pp. 415-417, 1998.
- [2.11] Jungwook Yang, Joongho Choi, Daniel M. Kuchta, Kevin G. Stawiasz, Petar Pepeljugoski, and H. A. Ainspan, "A 3.3-V, 500-Mb/s/ch parallel optical receiver in 1.2- μ m GaAs Technology," *IEEE J. Solid-State Circuits*, **33**, no. 12, pp. 2197-2204, 1998.
- [2.12] O. Vendier, N. M. Jokerst, and R. P. Leavitt "High efficiency thin-film GaAs-based MSM photodetectors," *Electronics Lett.*, **32**, no. 4, pp. 394-395, 1996.
- [2.13] N. V. Joshi, Photoconductivity, Marcel Dekker, New York, 1990.
- [2.14] Heinz Beneking, "Gain and bandwidth of fast near-infrared photodetectors: a comparison of diodes, phototransistors, and photoconductive devices," *IEEE Trans. Electron Devices* **ED-14**, pp. 1420-1430, 1982.
- [2.15] Stephen R. Forrest, "Optical detectors: three contenders," *IEEE Spectrum*, pp. 76-84; May, 1986.
- [2.16] M. DiDomenico, Jr., "Solid-state photodetection: a comparison between photodiodes and photoconductors," *Proc. IEEE*, pp. 136-144; February, 1963.
- [2.17] Richard A. Soref, "Silicon-based optoelectronics," *Proc. IEEE* **81**, pp. 1687-1706, 1993.
- [2.18] Frank Duncan, "Optimization of parameters in high-speed silicon photodiodes," *Electro-Optical Systems Design*, pp. 38-43; December, 1981.
- [3.1] S.M. Sze, D.J. Coleman, Jr., and A. Loya, "Current transport in metal-semiconductor-metal (MSM) structures," *Solid-State Electronics* **14**, pp. 1209-1218, 1971.

- [3.2] A. M. Cowley and S. M. Sze, "Surface states and barrier height of metal-semiconductor systems," *Journal of Applied Physics* **36** (10), pp. 3212-3220, 1965.
- [3.3] Y. C. Lim and R. A. Moore, "Properties of alternatively charged coplanar parallel strips by conformal mappings," *IEEE Trans. Electron. Device Lett.* **15**, pp. 173-180, 1968.
- [3.4] Wolfgang Hilberg, "From approximations to exact relations for characteristic impedances," *IEEE Transactions on Microwave Theory and Techniques* **17** (5), pp. 259-265, 1969.
- [3.5] Stephen Y. Chou, and Mark Y. Liu, "Nanoscale tera-Hertz metal-semiconductor-metal photodetectors," *IEEE J. Quantum Electronics* **28** (10), pp.2358-2368, 1992.
- [3.6] D.H. Auston, P. Lavallard, N. Sol, and D. Kaplan, "An amorphous silicon photodetector for picosecond pulses," *Appl. Phys. Lett.* **36** (1), pp. 66-68, 1980.
- [3.7] P.R. Smith, D.H. Auston, A.M. Johnson and W.M. Augustyniak, "Picosecond photoconductivity in radiation-damaged silicon-on-sapphire films," *Appl. Phys. Lett* **38** (1), pp. 47-50, 1981.
- [3.8] A. K. Sharma, K. A. M. Scott, S. R. J. Brueck, J. C. Zolper, and D. R. Myers, "Ion implantation enhanced metal-Si-metal photodetectors," *IEEE Photonics Tech. Lett.* **6** (5), pp. 635-638, 1994.
- [3.9] D.S. Shen, J.P. Conde, V. Chu, and S. Wagner, "Carrier lifetime in amorphous semiconductor," *J. Appl. Phys.* **75**, pp. 7349-7355, 1981.

- [3.10] D. R. Bowman, R. B. Hammond, and R. W. Dutton, "Polycrystalline-silicon integrated photoconductors for picosecond pulsing and gating," *IEEE Electron Device Letters* **EDL-6** (10), pp. 502-504, 1985.
- [3.11] R. B. Hammond, and N. M. Johnson, "Impulse photoconductance of thin-film polycrystalline silicon," *J. Appl. Phys.* **59** (9), pp. 3155-3159, 1986.
- [3.12] Massimo Ghioni, Franco Zappa, Vijay P. Kesan, and James Warnock, "A VLSI-compatible high-speed silicon photodetector for optical data link applications," *IEEE Transactions on Electron Devices* **43** (7), pp. 1054-1060, 1996.
- [3.13] Warren B. Jackson, N.M. Johnson and D.K. Biegelsen, "Density of gap states of silicon grain boundaries determined by optical absorption," *Appl. Phys. Lett.* **43** (2), pp. 195-197, 1983.
- [3.14] Peter W. Milonni, and Joseph H. Eberly, *Lasers*, Wiley Interscience, Toronto, 1988.
- [3.15] Ted Kamins, *Polycrystalline Silicon for Integrated Circuit Applications*, Kluwer Academic Publishers, Boston, 1988.
- [3.16] John Y.W. Seto, "The electrical properties of polycrystalline silicon films," *J. Appl. Phys.* **46** (12), pp. 5247-5254, 1975.
- [3.17] Y. Ishikawa, Y. Yamamoto, T. Hatayama, Y. Uraoka, and T. Fuyuki, "Quasi three-dimensional simulation for thin film polycrystalline silicon solar cells," in: *Conference Record of the Twenty-Eighth IEEE Photovoltaic Specialists Conference*, pp. 280-283, 2000.

- [3.18] Soumen Das and Samir K. Lahiri, "A large-bias conduction model of polycrystalline silicon films," *IEEE Transactions on Electron Devices* **41** (4), pp. 524-532, 1994.
- [3.19] N. C. C. Lu, L. Gerberg, C. Y. Lu, and J. D. Meindl, "A conduction model for semiconductor-grain boundary-semiconductor barriers in polycrystalline silicon films," *IEEE Trans. Electron Devices* **ED-30**, pp.137-149, 1983.
- [3.20] E. S. Yang, E. K. Poon, C. M. Wu, W. Hwang, and H. C. Card, "Majority carrier current chatacteristics in large-grain polycrystalline-silicon-Schottky-barrier solar cells," *IEEE Transactions on Electron Devices* **ED-28** (10), pp. 1131-1135, 1981.
- [3.21] G. Baccarani, M. Impronta, B. Ricco, and P. Ferla, "*I-V* characteristics of polycrystalline silicon resistors," *Revue de Physique Appliquee* **13**, pp. 777-782, 1978.
- [3.22] G. Baccarani, B. Ricco, and G. Spadini, "Transport properties of polycrystalline silicon films," *J. Applied Physics* **49** (11), pp. 5565-5570, 1978.
- [3.23] George J. Korsh and Richard S. Muller, "Conduction properties of lightly doped, polycrystalline silicon," *Solid-State Electronics* **21**, pp. 1045-1051, 1978.
- [3.24] H. C. Card and E. Yang, "Electronic processes at grain boundaries in polycrystalline semiconductors under optical illumination," *IEEE Transactions on Electron Devices* **ED-24**, pp. 397-402, 1977.
- [3.25] Dominique Ballutaud and Marc Aucouturier, "Electron spin resonance study of hydrogenation effects in polycrystalline silicon," *Applied Physics Lett.* **49** (23), pp. 1620-1622, 1986.

- [3.26] S. D. S. Malhi, H. Shichijo, S. K. Banerjee, R. Sundaresan, M. Elahy, G. P. Pollack, W. F. Richardson, A. H. Shah, L. R. Hite, R. H. Womack, P. K. Chatterjee, and H. W. Lam, "Characteristics and three-dimensional integration of MOSFETs in small-grain LPCVD polycrystalline silicon," *IEEE Trans. Electron Devices* **ED-32**, pp. 258-281, 1985.
- [3.27] I-Wei Wu, Tiao-Yuan Huang, W. B. Jackson, A. G. Lewis, and A. Chiang, "Passivation kinetics of two types of defects in polysilicon TFT by plasma hydrogenation," *IEEE Electron Device Lett.* **12** (4), pp. 181-183, 1991.
- [3.28] G. P. Pollack, W. F. Richardson, S. D. S. Malhi, T. Bonifield, H. Shichijo, S. Banerjee, M. Elahy, A. H. Shah, R. Womack, P. K. Chatterjee, "Hydrogen passivation of Polysilicon MOSFETs from a plasma nitride source," *IEEE Electron Device Lett.* **EDL-5**, pp. 468-470, 1984.
- [3.29] R. T. Young, M. C. Lu, R. D. Westbrook, and G. E. Jellison Jr., "Effect of lithium on the electrical properties of grain boundaries in silicon," *Applied Physics Lett.* **38** (8), pp. 628-630, 1981.
- [3.30] E. Puppini, "Hydrogen implanted polycrystalline silicon: resistivity and grain boundary chemistry," *J. Vacuum Science Technology B* **5**, pp. 606-607, 1987.
- [3.31] T. I. Kamins, and P. J. Marcoux, "Hydrogenation of transistors fabricated in polycrystalline-silicon films," *IEEE Electron Device Lett.* **EDL-1**, pp. 159-161, 1980.
- [3.32] C. H. Seager and D. S. Ginley, "Passivation of grain boundaries in polycrystalline silicon," *Appl. Phys. Lett.* **34** (5), pp. 337-340, 1979.

- [3.33] T. Makino and H. Nakamura, "The influence of plasma annealing on electrical properties of polycrystalline Si," *Appl. Phys. Lett.* **35** (7), pp. 551-552, 1979.
- [3.34] D. R. Campbell, "Enhanced conductivity in plasma-hydrogenated polysilicon films," *Appl. Phys. Lett.* **36** (7), pp. 604-606, 1980.
- [3.35] Theodore I. Kamins and Richard S. Muller, Device Electronics for Integrated Circuits, John Wiley and Sons, Toronto, 1986.
- [3.36] T. I. Kamins, "Hall mobility in chemically deposited polycrystalline silicon," *J. Appl. Physics* **42** (11), pp. 4357-4365, 1971.
- [3.37] K. Shibata, Y. Ohmura, T. Inoue, K. Kato, Y. Horiike, and M. Kashiwagi, "Growth of large grained silicon on insulator by electron beam annealing and performance of MOS devices," *Japanese J. Appl. Physics* **22**, Supplement 22-1, pp. 213-216, 1983.
- [3.38] M. Hack and A. G. Lewis, "Avalanche-induced effects in polysilicon thin-film transistors," *IEEE Electron Device Lett.* **12** (5), pp. 203-205, 1991.
- [4.1] R.G. Smith and S.D. Personick, "Receiver design for optical fiber communication systems," in Semiconductor Devices for Optical Communication, H. Kressel, Ed. Springer-Verlag, New York, 1980.
- [4.2] Wei Gao, Paul R. Berger, R. G. Hunsperger, G. Zydzik, W. W. Rhodes, H. M. O'Bryan, D. Sivco, and A. Y. Cho, "InGaAs MSM photodiodes with transparent Schottky contacts and Digital superlattice grading," in: 15th North American Conference on Molecular Beam Epitaxy, College Park, MD, U.S.A., 1995.
- [4.3] Nan M. Jokerst, Martin A. Brooke, J. Laskar, D. Scott Wills, A. S. Brown, M. Vrazel, S. Jung, Y. Joo, and J. J. Chang, "Microsystem optoelectronic integration

- for mixed multisignal systems,” *IEEE J. Sel. Topics Quantum Electronics*, **6**, no.6, pp.1231-1238, 2000.
- [4.4] J. Y. L. Ho and K. S. Wong, “Bandwidth enhancement in silicon metal-semiconductor-metal photodetector by trench formation,” *IEEE Photon. Technol. Lett.* **8**, pp. 1064-1066, 1996.
- [4.5] Medici, Two-dimensional device simulation program, version 4.0, Technology Modeling Associates, Inc., 1997
- [4.6] Douglas M. Considine, editor, Van Nostrand’s Scientific Encyclopedia, Van Nostrand Reinhold Company, Toronto, 1976.
- [4.7] Paul R. Berger, “MSM photodiodes,” *IEEE Potentials*, pp. 25-29, April/May, 1996.
- [4.8] M. Klingenstein and J. Kuhl, “Photocurrent gain mechanisms in metal-semiconductor-metal photodetectors,” *Solid-State Electronics* **37**, pp. 333-340, 1994.
- [4.9] R. J. Schreutelkamp and L. Deferm, “A new method for measuring the saturation velocity of submicron CMOS transistors,” *Solid-State Electronics* **38** (4), pp. 791-793, 1995.
- [4.10] S. Alexandrou, Chia-Chi Wang, T. Y. Hsiang, M. Y. Liu, and S. Y. Chou, “A 75 GHz silicon metal-semiconductor-metal Schottky photodiode,” *Appl. Phys. Lett.* **62**, pp. 2507-2509, 1993.
- [4.11] D. L. Rogers, “Interdigitated metal-semiconductor-metal detectors,” *Proc. Picosecond Electronics and Optoelectronics* **4**, pp 184-188, 1989.

- [4.12] J. B. D. Soole and H. Schumacher, "InGaAs metal-semiconductor-metal photodetectors for long wavelength optical communications," *IEEE J. Quantum Electronics* **27**, pp. 737-752, 1991.
- [4.13] W. Roth, H. Schumacher, J. Kluge, H. J. Geelen, and H. Beneking, "The DSI diode – a fast large-area optoelectronic detector," *IEEE Trans. Electron Dev.* **ED-32**, pp 1034-1036, 1985.
- [4.14] H. Beneking, "On the response behavior of fast photoconductive optical planar and coaxial semiconductor detectors," *IEEE Trans. Electron Dev.* **ED-29**, pp 1431-1441, 1982.
- [4.15] John Bardeen, "Surface states and rectification at a metal semi-conductor contact," *Physical Review* **71**, pp. 717-727, 1947.
- [4.16] W. A. Wohlmuth, P. Fay, and I. Adesida, "Dark current suppression in GaAs metal-semiconductor-metal photodetectors," *IEEE Photonics Tech. Lett.* **8** (8), pp. 1061-1063, 1996.
- [4.17] J. C. Bean, Jieming Qi, C. L. Schow, R. Li, H. Nie, J. Schaub, and J. C. Campbell, "High-speed polysilicon resonant-cavity photodiode with SiO₂-Si Bragg reflectors," *IEEE Photonics Tech. Lett.* **9** (6), pp. 806-808, 1997.
- [4.18] D. Kuhl, F. Hieronymi, E. H. Bottcher, T. Wolf, D. Bimberg, J. Kuhl, and M. Klingenstein, "Influence of space charges on the impulse response of InGaAs metal-semiconductor-metal photodetectors," *J. Lightwave Tech.* **10** (6), pp. 753-759, 1992.

- [4.19] F. Hieronymi, E. H. Bottcher, E. Droge, D. Kuhl, and D. Bimberg, "High-performance large-area InGaAs metal-semiconductor-metal photodetectors," *IEEE Photonics Tech. Lett.* **5** (8), pp. 910-913, 1996.
- [4.20] Masanori Ito, and Osamu Wada, "Low dark current GaAs metal-semiconductor-metal (MSM) photodiodes using WSi_x contacts," *IEEE J. Quantum Electronics* **QE-22** (7), pp. 1073-1077, 1986.
- [4.21] R. G. DeCorby, R. I. MacDonald, R. Sharma, F. Gouin, J. Noad, and N. Puetz, "Frequency-domain demonstration of transit-time-limited, large-area InGaP-InP-InGaAs MSM photodetectors," *IEEE Photonics Tech. Lett.* **9** (7), pp. 985-987, 1997.
- [4.22] D. Kuhl, F. Hieronymi, E. H. Bottcher, and D. Bimberg, "High-speed metal-semiconductor-metal photodetectors on InP:Fe," *IEEE Photonics Tech. Lett.* **2** (8), pp. 574-576, 1990.
- [4.23] B. Jalali, L. Naval, and A. F. J. Levi, "Si-based receivers for optical data links," *J. Lightwave Tech.* **12** (6), pp. 930-934, 1994.
- [4.24] Takayuki Sugeta, Tsuneo Urisu, Seizo Sakata, and Yoshihiko Mizushima, "Metal-semiconductor-metal photodetector for high-speed optoelectronic circuits," *Japanese Journal of Applied Physics* **19**, supplement 19-1, pp. 459-464, 1980.
- [5.1] R. Hadaway, T. Brown, K. Harris, R. Foucault, "BiCMOS technology for telecommunications," IEEE 1993 Bipolar Circuits and Technology Meeting, pp. 159-166, (1993).
- [5.2] Sherif H.K. Embabi, Abdellatif Bellaouar and Mohamed I. Elmasry, Digital BiCMOS Integrated Circuit Design, Kluwer Academic Publishers, Boston, 1993.

- [5.3] Paul E. Green, Jr., Fiber Optic Networks, Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [5.4] J. J. Morikuni, A. Dharchoudhury, Y. Leblebici, and S. M. Kang, "Improvements to the standard theory for Photoreceiver noise," *J. Lightwave Tech.* **12** (4), pp. 1174-1183, 1994.
- [5.5] S. D. Personick, "Receiver design for digital fiber optic communication systems, Parts I and II," *Bell Syst. Tech. J.* **52** (6), pp. 843-886, 1973.
- [5.6] Alan B. Grebene, Bipolar and MOS Analog Integrated Circuit Design, John Wiley & Sons, Toronto, 1984.
- [5.7] Hans-Martin Rein, "Silicon bipolar integrated circuits for multigigabit-per-second lightwave communications," *J. Lightwave Tech.* **8** (9), pp. 1371-1378, 1990.
- [5.8] A.V. Krishnamoorthy et. al., "3-D integration of MQW modulators over active submicron CMOS circuits: 375 Mb/s transimpedance receiver-transmitter circuit," *IEEE Phot. Tech. Lett.* **7** (11), pp. 1288-1290, 1995.
- [5.9] Reinhard Reimann and Hans-Martin Rein, "Bipolar high-gain limiting amplifier IC for optical-fiber receivers operating up to 4 Gbit/s," *IEEE J. Solid-State Circuits* **22** (4), pp. 504-510, 1987.
- [5.10] P. Myslinski, "Ultra-high-speed photodetectors," NRC internal report, contract no. 990-1122/1706/1990, January 1991.
- [5.11] H.-M. Rein, M. Neuhauser, H. Wernz, V. Flassnocker, "A 6.5 Gbit/s transimpedance preamplifier in silicon bipolar technology for optical-fibre transmission links," *Frequenz* **46** (5-6), pp.174-176, 1992.

- [5.12] Ronald A. Nordin and Anthony F.J. Levi, "A system design perspective on optical interconnection technology," *AT&T Tech. J.*, pp. 37-49, September/October 1993.
- [5.13] J. Wieland et al., "Optical receiver array in silicon bipolar technology with selfaligned, low parasitic III/V detectors for DC-1 Gbit/s parallel links," *Electron. Lett.* **27** (24), pp. 2211-2213, 1991.
- [5.14] R.L. Calton et al., "The design and performance of the PTAT-1 optical receiver," *Telecom. Tech. J.* **7** (1), pp. 78-82, 1989.
- [5.15] Tran Van Muoi, "Receiver design for high-speed optical-fiber systems," *J. Lightwave Tech.* **2** (3), pp. 243-267, 1984.
- [6.1] Stanley Wolf, "Silicon Processing for the VLSI Era, Volume 1: Process Technology" Lattice Press, Sunset Beach, U.S.A., 1986.
- [6.2] Sorab K. Ghandhi, "VLSI Fabrication Principles," John Wiley and Sons, Toronto, 1983.
- [6.3] Stanley Wolf, "Silicon Processing for the VLSI Era, Volume 2: Process Integration" Lattice Press, Sunset Beach, California, 1990.
- [6.4] A. Brand, A. Haranahalli, N. Hsieh, Y.C. Lin, G. Sery, N. Stenton, B.J. Woo, S. Ahmed, M. Bohr, S. Thompson, S. Yang, "Intel's 0.25 Micron, 2.0Volts Logic Process Technology," *Intel Technology Journal*, Q3 1998,
<http://www.intel.dk/technology/itj/q31998.htm>.
- [6.5] Ben G. Streetman, "Solid State Electronic Devices," Prentice Hall, Englewood Cliffs, New Jersey, 1990.