Electro-optic Platform for Free Space CMOS Photonics.

by

Serb, Alexantrou

October 2013

A thesis submitted for
the degree of Doctor of Philosophy of Imperial College

Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine

Contents

List of Figures

List of Tables

Abstract

Lab-on-chip (LOC) systems are becoming increasingly popular for biomedical science as they present the opportunity to combine compact and efficient microelectronics together with microfluidics enabling new applications in point-of-care diagnostics. This system integration however, poses significant challenges in the assembly of such devices for mass manufacture. Specifically, to achieve a robust fluidic isolation, an insulating material must be deposited to seal the chip and wire bonds but allow fluid to access the sensing surface. This is typically achieved by using an insulating epoxy for encapsulation but requires several processing steps in order to become planar and reliably interface with the microfluidics; a technique with many limitations.

Towards addressing these challenges, this thesis proposes to develop a non-galvanic means of achieving both power transfer to the chip and bi-directional data communication such that the system requires no bond pads (or delicate bond wires). The aim is to achieve this specifically via a free space optical link (i.e. to external discrete optoelectronic devices) with the additional constraint that any structures designed on chip are implementable in a commercially available, unmodified CMOS technology. Furthermore, in order to maintain the desired benefits of a bondpad-less chip, the platform must utilise no off-chip components. This thesis develops the underlying devices required towards achieving the aim whilst satisfying all constraints. Specifically devices tasked with optical energy harvesting, optical data input and optical data output are tackled.

The thesis begins by outlining the motivations for this research (Chapter 1) and reviewing the relevant state-of-the-art (Chapter 2) including a concise overview of alternative methods for achieving the underlying aims. The relevant theory, pertinent to electro-optical phenomena at semiconductor junctions is then developed within the context of CMOS technology (Chapter 3). More fundamental, background theory is also included in appendix A. That pertains to the propagation of light in Silicon and mechanisms of photon absorption in doped Silicon. Original contributions within the domain of theory include developing the phenomenon of free-carrier absorption (FCA) applied to a realistic, CMOS-based junction, identifying key variables, expressions and analyzing the expected level of performance.

Then, for the first time, this thesis demonstrates free-space optical modulation in a standard CMOS technology. A large portion of this work (Chapter 4 - for design repository see appendix B) is thus devoted to the design, implementation and testing of prototype devices for use as data read-out elements. A variety of modulator devices featuring differing geometries, created by distinct doping procedures and implemented in different CMOS technological nodes (UMC 0.13 μm, IBM 0.18 μm and AMS 0.35 μm) are presented, tested and compared. This allows for modulator performance to be examined in relation to key design choices made at the physical device layout and technology choice levels.

This thesis then develops the common data read-in and power scavenging mechanism, along with associated circuits (Chapter 5). Once again structures designed with different geometries and created by different manufacturing processes in different technological nodes are presented, tested and compared, yielding an indication towards underlying trends. Key contributions here include extracted photodiode

model parameters that express the contributions of vertical and lateral junction components to photocurrent generation, by junction 'family' and by CMOS process. This provides a powerful resource to circuit designers requiring a first estimate to phototransduction efficiency in technologies with unspecified optoelectronic devices.

Declaration of originality

I hereby declare that the work contained within this thesis is my own. The utmost effort has been committed to ensure that any external content has been properly referenced.

Copyright declaration

Acknowledgements

This work could not have been completed without the help of the following people:

Supervision: Drs. Timothy Constandinou and Konstantin Nikolic have given me all the support I have and indeed could have asked for throughout this project. They guided me from beginning to end and helped me mature both as a scientist and as a person. To them I owe a great debt.

Technical support: Mrs. Khaleeq Maria has given me great help with great patience throughout the years. Bonding the unbondable and mounting the unmountable, she made sure that all the dies described in this thesis are properly packaged. moreover, credit should also go to Mr. Phil Jones who kindly provided the chip packages used with clean and well-centred $3\,mm$ diameter holes.

General support: There are many people who have stood by me throughout the years of my PhD and given me the pleasure of exciting conversations, or have kindly helped me when I found myself in need of their abilities. My heart-felt thanks go to: Melina Kalofonou, Song Luan, Yan Liu, Themistoklis Prodromakis and many others.

Personal support: Thanks go to my parents, grandparents and my housemates for their ongoing support during what has not always been smooth sailing.

Companies: Finally, many thanks to our suppliers who have performed their tasks without problems. Particular thanks to Europractice who have provided us support throughout what every designer knows are always chaotic times: tape-out seasons.

Chapter 1

Introduction

1.1 Motivation and Context

In the electronics industry wire-bonding is a well-known step which is often discarded as a mere technicality. However no die is useful unless it has the means to communicate with the outside world and receive power. In most cases what one might call the 'three basic functions' of any link between chip and board, namely: data input, data output and power input, are achieved by means of wire-bonds and as such the importance of these minute wires should not be underestimated.

In electronic component mass-production there are automated systems that can handle high volumes of wire-bonding thanks to the standardisation of components and the well-established methods used to manufacture plastic or sometimes ceramic packaging. Lower volume applications typically have to rely on a technician who will carry out the task manually with the same applying for 'specialist jobs' whereby the wire-bonding procedure is not merely a matter of connecting an integrated bond-pad to the pad of a package, but needs to comply to other specifications as well.

Perhaps the best example of such situation is when a die includes chemical sensors or any sort of microfluidic element. For such applications the bond-wires can simply not be allowed to connect the die to the package through free air because of the implications this will have once the die and its surrounding area is flooded by the test liquid. The conductivity of the liquid will cause shorts between the bond-wires, which in itself is reason enough to take action. Typically when a die has to come in close proximity with liquids the packaging solution involves depositing insulating epoxy on top of the wire-bond with extreme care. This is because should the epoxy not fall and settle within a very accurately defined area either sensitive parts of the die are going to be buried underneath or bond-wires are going to be left bereft of any protection against the deluge of liquid. This will either compromise the function of the die or unduly short bond-wires amongst themselves. To compound these problems an extra step might need to be added to the design flow if the final product is meant to be planar (see Figure 1.1).

Clearly the complications that arise from having to meticulously tend to the bond-wires are a matter that could use a solution. To that end developing a contactless power and data transfer interface would be desirable. Within this domain a couple of methods stand out as potential candidates for practical implementation on an industrial scale: design centred around inductive coupling and design based on optoelectronics.

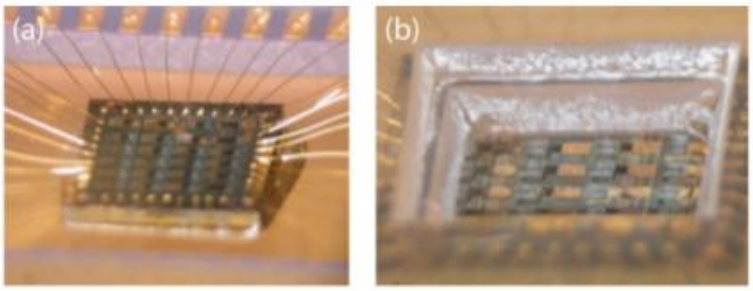

Figure 1.1: Traditional packaging and encapsulation of lab-on-chip systems. Shown are: (a) traditional encapsulation whereby the entire die with all bond wires (shown exposed in the picture) all end up buried deep inside the (usually plastic) casing. (b) how an experimental set-up must be altered to ensure the functional coexistence of fluidics and electronics. A layer of insulating epoxy has been added so that is covers all the bond wires, but leaves the surface of the die exposed.

Inductive coupling suffers from a couple of major issues, namely poor relationship between efficiency and distance and severe cross-talk. Both these problems arise from the properties of the EM field that can be generated by an inductive coil, which decreases dramatically with distance from the coil even along the principal axis that runs through the centre of the cylindrical coil and diverges equally dramatically with distance. Moreover, because of its electromagnetic nature the cross-talk doesn't only affect adjacent channels of inductor-based communication but will also create noise on every piece of conductor serving the circuitry lying nearby. Nevertheless, if cross-channel communication is assumed not to be a problem (as in the case of a single-channel system) and the power & data transmission is carried out at a frequency chosen not to interfere with the function of the circuit the system can theoretically work rather well, but the cost is rather stringent design constraints. These considerations will be discussed in further detail in the next chapter.

Optoelectronic methods offer the other realistic alternative when it comes to wireless chip to board coupling. Optical power scavenging is a well-established field with a lot of research going into manufacturing efficient photovoltaic elements. Integrated circuits have not escaped this trend with work done on creating efficient power recovery structures in dies of various substrate materials and technological nodes. Interestingly the same principles used for power scavenging from optical sources can be easily adapted for use towards capturing external optical signals. The real problem, however, is the matter of recovering a signal generated on-chip. For that there are only two possible routes: either the die is somehow made to emit light that can be captured by an external detector, or a steady beam can be sent through the die, modulated somewhere therein and then captured by an external detector. Tackling this issue forms the crux of this PhD thesis and is considered extensively in following chapters. Solving this problem opens the door to leveraging the benefits of the optical way of creating a wireless chip to board communication platform: the fact that beams can be formed so that they diverge minimally (by use of collimators for example) and that signals communicated back and forth through this interface are optical in nature which means that they should not interfere too much with circuitry that happens to lie nearby, thus granting access to a very 'clean' way of passing information and power to the die.

1.2 Towards a fully contact-less optoelectronic platform

The focus of this PhD thesis is implementing an optoelectronic, bi-directional data and power transfer platform for use in a general context, but with specific focus to the needs of the microfluidics branch of science. As such the main priority is to get rid of the need for bond-wires altogether, even if it comes at the expense of speed, and ultimately produce demonstrator devices that can successfully perform the aforementioned 'three basic functions' without having to rely on any bond-wires.

Another constraint that comes into play is the possibility of implementing this technology in commercially available CMOS technology. Tailoring a die to incorporate simple optoelectronic devices such as photodiodes and integrated light emitters becomes much easier if specialised manufacturing technologies are called into action. Elements such as dislocation loops or embedded hetero-structures provide a solution to the otherwise impossible problem of coaxing an indirect bandgap like silicon into emitting light with high degree of efficiency. Nevertheless, most integrated circuit design is carried out in fairly simple CMOS or BiCMOS technologies that offer little more than a selection of Boron-, Phosphorus- and Antimony-doped diffusion and well regions. As none of these structures are designed to or indeed manage to turn silicon into an efficient light emitter the matter of fact is that most IC design is performed without any inherent support for light emitting structures. For that reason creating an optoelectronic communications platform in commercially available CMOS technologies introduces an extra challenge but also renders the technology a lot more usable in a practical setting.

Given the constraints discussed above key elements of a viable optoelectronic communications platform begin to emerge. To begin with, a number of photo-sensitive structures will be needed in order to harvest power from an external optical power source whilst a set of separate but probably closely related structures could be used to receive an optical input signal. These structures are nothing conceptually new, however keeping them separate from one another and ensuring low cross-talk in a practical setting will involve the use of precision optics. On the flip side data outflux would have to be carried out by modulator structures. This strategy is chosen because silicon may be a bad light emitter and at the same time not a brilliant modulator, but in the case of modulation this inherent weakness of silicon does not lead to large power dissipation in the effort to compensate for it. Finally a viable power management unit needs to be introduced in order to upregulate and stabilise the power arriving from the power scavenging elements at a relatively low voltage. The conceptual block diagram in Figure 1.2 illustrates these constituent parts of the system, as well as how they fit together and work to service the circuits residing on the die.

Achieving light modulation in Silicon is not a trivial task, but it has been done before as we shall see in the literature review chapter (Chapter 2). Modulation can typically be achieved in terms of electro-refraction (Kerr effect, Free Carrier refraction for example) or electro-absorption (Franz-Keldysh effect, Free Carrier absorption).

As an example, if one is to use the free carrier absorption phenomenon as the vehicle for modulation then all that is required is a reverse-biased pn-junction to fulfill the role of modulator and enough energy to charge and discharge the junction capacitance when switching digital state. This is particularly attractive because modulating the reverse bias at a micro-electronic pn-junction is known not to be an energetically taxing activity. Furthermore, complicated, coherent light sources, interferometers to detect phase shifts and tight optical path control are all not required if the modulated quantity is light

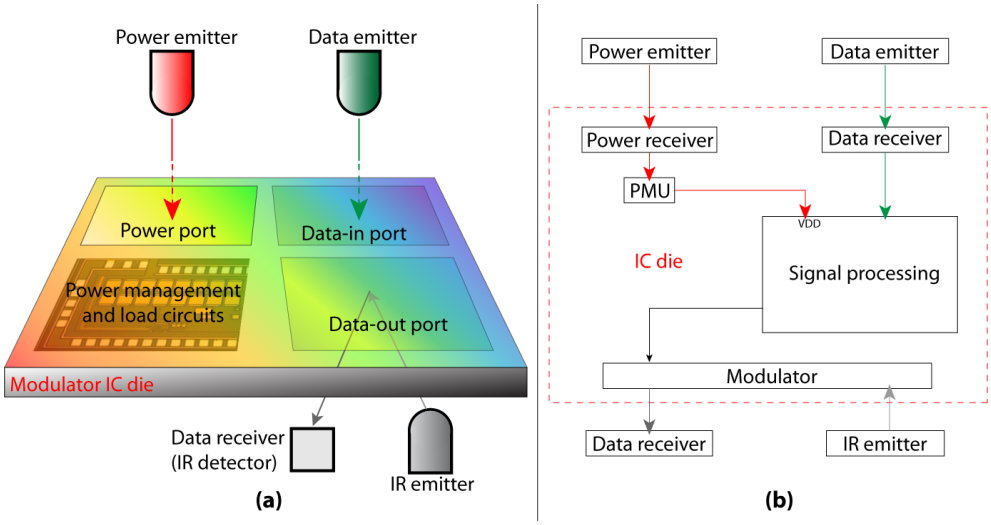

Figure 1.2: Conceptual power/data port configuration of a contactless chip. The physical manifestation is shown in panel (a), where three generic pn-junctions are shown in the roles of power receiver, data receiver and modulator. The block diagram is shown in panel (b) where a bit more detail of the internal structure of the contactless IC is shown. Power pathway: red. Inbound signal pathway: Green. Modulatory signal pathway: Gray (optical) and black (electrical).

amplitude. This is important in our set-up because our optical path is very loosely controlled and subject to serious manufacturing variation. Thus, it is mainly for these reasons, that the engineering of the Free Carrier Absorption (FCA) phenomenon is the chosen technique for achieving modulation in this project.

On the other hand, Free Carrier Refraction has been used in waveguide-based systems that have been extensively developed over the past decades (see chapter on literature review -Chapter 2- for full discussion). It can be quite useful and perform as very high data transfer speeds (as we shall see in the literature review chapter) when very high quality and tightly controlled coherent light sources exist but because typically the effects used revolve around phase modulation and resonance in micro-cavities the effect is hard to engineer with less expensive light sources and less precision-engineered optical paths.

In summary, the main contributions of this thesis lie in the development and testing of devices capable of exploiting optical phenomena in order to allow contact-less communication with an integrated circuit by purely optical means: electro-optical modulators for data read-out and optical power capture cells for data read-in and power recovery. Secondary contributions were made along the way towards achieving these goals, primarily in the domains of integrated circuit packaging and theory pertaining to the study of our resulting electro-optical devices.

1.3 Thesis outline

The thesis henceforth is organised as follows:

- Chapter 2 shall provide a review of the state of the art and associated literature. This will include papers tackling the issues of power recovery in CMOS chips, the design of power management units for scavenging chips and provide some work done on the free carrier absorption phenomenon

that shall put our decided mechanism of modulation into perspective. A close examination of the method of creating contact-less chips via inductive coupling will also be performed in this chapter.

- Chapter 4 shall present the technical/theoretical basis of the modulator that will allow the die to send data to the outside world. Practical considerations that mark the difference between theory and practice are part of this chapter.

- Chapter 3 pertains to background theoretical information. The physics of the phenomenon will be considered in this section. Aside from that, physics related to photo-transduction and basic theory relating to pn-junctions and ideal semiconductor doing processes are also included in these chapters; all important elements in understanding the full system and placing it into perspective.

- Chapter 5 forms the core of the thesis as information about results obtained from our modulator devices is included here. The chapter covers subjects ranging from modulator design, through the result presentation and processing, to result interpretation and additional considerations.

- Chapter 6 will detail the characterisation and comparison of power scavenging structures including photodiodes manufactured in different technologies, using different types of pn-junctions and with different geometrical characteristics. A discussion on a sample power management unit (PMU) used for demonstrative purposes will also be included in this chapter.

- Chapter 7 will conclude this thesis with a summary of the work presented throughout and a brief, overarching discussion of the results and their significance in terms of pure science and applications. Moreover, an indication will be given as to how this research may be taken to the next step.

- Appendix A will be presenting various designs of integrated modulators produced during the length of this research. Essentially this chapter constitutes the 'modulator design library' where information on the vital statistics of each modulator and the motivation behind each individual design is stored. Note: our modulator devices are also used for testing in their power harvesting capabilities so information here relates to the chapters regarding power harvesting as well.

- Appendix 5.4 includes information on the test-bench used to extract data related to electro-optical modulators.

- Appendix B contains raw data measured from experiments designed to understand electro-optical modulation in our test structures.

Chapter 2

Contactless Chips: State-of-the-Art

This chapter is intended to review the literature in the domain of contactless integrated circuits. The core of the chapter pertains to semiconductor structures and integrated systems that can achieve the three basic functions of a contactless system: send data, receive data and scavenge enough power to keep the system operational. As such, data and power read-in can be achieved by the use of photodetector elements whilst data read-out can achieved by either emitting light straight from the die or by modulating a beam that passes through the light. Light emitting, modulating and capturing structures are all reviewed in this chapter both as individual structures and within the context of full data/power transceiver systems.

Unfortunately, the nature of optoelectronic elements means that particularly when it comes to power scavenging their interface with any useful circuit (load circuit for power supplies, signal processing circuits for signalling elements) is rather difficult to achieve with success. For example, special circuitry must transform the power output of the actual harvester devices into a form that can be realistically used by load circuits. This implies restrictions as to how low the voltage can be and how intensely it can be allowed to fluctuate. For that reason power management circuits are used to convert the power output of scavenger devices from an excessively low and unstable 'inconvenient' form into a sufficiently high voltage and stable 'suitable' form. Practically every integrated circuit that powers itself by means of optical power scavenging requires such a unit since power supply devices can hardly supply more than 0.7 V (when implemented in Silicon), a voltage headroom that is insufficient for too many applications. Therefore a section of the literature review will be devoted to power management circuits. On the other hand, signal read-in and signal read-out can conceivably be achieved easily provided that a sufficiently powerful power supply with enough headroom is available; a task that falls upon the power management unit to achieve.

Other literature that shall be reviewed within this thesis concerns the related topics of inductively coupled systems and non-contactless packaging. Inductive coupling will be reviewed because it too can achieve contactless bidirectional power and data transmission. It is, thus, a direct competitor to the optical coupling method. Finally, non-contactless packaging will also be discussed as the sets of strengths and weaknesses its various forms come with will set the scene for discussing its contactless alternatives.

In summary: The journey throughout the related literature begins with an overview of non-contactless microelectronic packaging (Section 2.1). Next, the achievement of a contactless power and data communications platform via inductive coupling methods is discussed (Section 2.2). Silicon photonic structures

that could be potentially used to implement an optoelectronic system capable of achieving the aforementioned bidirectional data and power transfer are then reviewed (Section 2.3). Supporting circuitry that appears in power management systems for optical energy scavengers is considered next (Section 2.4) before optical transceivers featuring the structures reviewed in Sec. 2.3 and the circuits in Sec. 2.4 are discussed within the context of full systems (Section 2.5). Finally a summary (Section 2.6) marks the end of the literature review chapter.

2.1 CMOS lab-on-chips: Packaging & encapsulation challenges

The idea of integrating various electronic modules on a single chip in order to perform tests that would otherwise require large numbers of external instruments is not new. However, when it comes to creating such 'lab-on-chip' (LOC) systems for microfluidic applications the issue of first encapsulating and then planarising the bond wires is also not new. Subsequent problems pertaining to the quality of the seal are also not uncommon. First, we shall examine the mainstream techniques for chip encapsulation, then we shall examine an array of newer and more advanced techniques developed specifically to improve packaging for dies that have to come into contact with liquid elements, and finally we shall summarise the literature on the issue of chip encapsulation and mention a few more papers that have dealt with the issue over time.

2.1.1 Standard encapsulation techniques

The vast majority of dies are encapsulated either through transfer moulding, potting or glob-top processes[1]. These are shortly reviewed in [1] but a brief description is repeated here in the interests of the reader.

2.1.1.1 Transfer moulding

Transfer moulding relies on the concept that the die is encased into a heated mould which is subsequently filled with moulding material under high pressure. Conventional integrated circuits are often encapsulated using this method because of its high speed [1]. A variant of the process has been developed and patented [2] that performs selective encapsulation, thereby allowing direct access to parts of the surface of the die after processing, i.e. it leaves open 'windows' on the top side of the package. This is achieved by simply ensuring that the mould is constructed in such way as to sit softly, but tightly over the parts of the die that we wish to leave exposed. As such when the mould is applied on the die + package, encapsulant fills all critical parts, but is denied access to a significant portion of the die surface, thus creating the aforementioned 'windows' of exposed die. A version of this technique described from a more academic (as opposed to industrial) perspective can be seen in this late 90's work [3]. After 2000 this technique continued to improve, notably with the advent of Film Assisted Molding (FAM), as detailed in this paper [4]. The innovation here is the application of the thin film is used to improve the quality of the moulding and the longevity of the actual mould whilst the basic process behind it remains the same. Meanwhile, in the same decade the industry has also developed a highly application-specific method of

[1] Other methods that exist, such hermetic metal and glass case encapsulations would not appear to be well-suited for microfluidic applications and are not studied in the present work.

selectively encapsulating parts of dies (the parts that host the wire-bonds) and applied this technique for manufacturing chemical sensing instrumentation based on chips carrying ISFETs (Ion-Sensitive Field Effect Transistors). A very detailed study of various specific ways of encapsulating ISFET-carrying dies is provided in this 2005 paper [5] along with information on the limitations that these packages impose in the field of ISFET chemical sensing. Finally, another interesting development in mould-based encapsulation has been described [[6]] where the group has used flexible PCBs to connect a die that is tucked into an alcove within the main, rigid PCB to the main PCB and then applied a top-mould over the complex and filled the space between with encapsulant. Windows were left for the fluid to be able to access the chemical-sensitive parts of the die. Although issues with alignment and side-wall isolation may prove challenging to overcome in the lab, the process produces undoubtedly slimmer resulting assemblies and eliminates wire-bonds altogether in favour of a flexible PCB-based packaging approach.

2.1.1.2 Potting

Potting techniques are simpler than transfer moulding and revolve around the concept that the die should be glued and wire-bonded within a cavity which is subsequently flooded with encapsulant and then left to cure. Naturally the process involves extra steps to make sure that no air-filled cavities are left. Though the list of companies that offer potting services is long, academic literature on the subject is rather modest with much having been said about this process in a handful of books, e.g. [7] and very little elsewhere. The author could not find any mentions of any techniques related to potting used to selectively encapsulate a die even though lowering a metallic rod with a custom plastic tip ('stopper tip') into the pot cavity until it touches the surface of the die and then releasing the encapsulant would not seem as a particularly difficult task to do. Naturally practical considerations such as those pertaining to alignment of the stopper tip with the die, the choice of materials for the tip, the amount of pressure applied on the stopper etc. would all need to be considered were such method to find its way into commercial application.

2.1.1.3 Glob-top encapsulation

Glob-top processes could be described as dropping blobs of liquid encapsulant over the die and then curing it. Though this method appears very straightforward, particularly for cases where the die must be encapsulated completely and where an encapsulant that cures at room temperature is used, extra processing steps may be needed in order to render the technique suitable for partial encapsulations. For that reason, when the time comes to deal with a chip that needs to have its bond wires protected but at the same time requires that part of it remains exposed, dual resin processes can be employed. A high viscosity 1st resin is applied to begin with and the aim is to delimit the areas which will be later on flooded by the less viscous 2nd resin. This technique is known as 'damming' because the 1st encapsulant forms 'dams' which are used to contain its 2nd counterpart. The previously referenced thesis [1] describes a method whereby damming is used to selectively encapsulate the die. According to this technique, a double dam is formed: one outer wall to prevent spillage of the 2nd encapsulant into the ambient environment and an inner wall that 'walls out' the sections of the die that must be kept exposed after packaging. The resulting cavity between the inner and outer walls is filled with the 2nd encapsulant and the set-up is allowed to cure. Once again there might be a case for using a metallic rod with a suitable stopper tip in order to protect the parts of the die that need to stay exposed instead of

resorting to using an inner wall, but such technique has not been mentioned in the literature to the best knowledge of the author.

The three standard methods are illustrated in Figure 2.1.

2.1.2 Advanced encapsulation methods

A number of advanced encapsulation techniques have been proposed in the literature in an attempt to create better packaging for 'specialist applications', notably microfluidic applications. Creating the microfluidic channel at the same time as encapsulating the target die is no trivial task. A number of proposed solutions are examined below.

2.1.2.1 Exploitation of capillarity action

Amongst the more advanced encapsulation techniques there is one that is described in this 2006 paper [8]. The group here used UV-curing encapsulants and exploited capilarity forces in order to ensure that glue floods a predetermined area at the interface between the microfluidic channel component and the board. The hydrophilicity of board, channel and glue had to be matched for this to succeed. This follows earlier work on the idea of exploiting capillarity action for the purposes of die encapsulation [9].

2.1.2.2 Photolithographic methods

Meanwhile a 2008 paper [10] discusses the merits of a photolithographic approach to encapsulation using photoresist to pattern the encapsulant and then etch away material from the areas where the die must ultimately be exposed. Later on, in 2010, this paper [11] uses a similar concept to successfully encapsulate a CMOS chip with parylene. However, relying on capillarity to draw encapsulant and then curing it via UV at exactly the right moment is not a trivial process even though it may produce perfectly valid results. On the other hand the use of photolithographic techniques to pattern an encapsulant involves the use of rather complicated and expensive equipment and from what results from the above-stated literature the encapsulation must then be completed at wafer level because of the limitations of the spin-coating process[2]. This implies coating the entire PCB at once which, amongst others, places restrictions upon the nature of the PCB itself. After all, spin-coating a single die is hardly an easy process.

2.1.2.3 mm-scale gasketing

In another attempt to solve the encapsulation conundrum that plagues microfluidic chips, Rothberg and colleagues [12] discusses an interesting gasket-based encapsulation technique for larger dies (on the order of cm^2) whereby a gasketed structure with a fluidic inlet and outlet is mechanically pressed against the die, thus creating an effectively sealed chamber over the centre of the chip. The wire-bonds stay outside this gasket, of course. This method is simple and quite efficient (the gasket can always be replaced in case of failure), but seems to work for rather sizeable dies only. Specifically the dies under consideration

[2]Semiconductor fabrication procedure whereby a surface is coated by a thin, even film of material. The procedure works thus: Initially, a blob of coating material is applied to the centre of a surface to be coated (typically a roughly circular wafer). The surface is then rotated under controlled conditions (temperature, angular velocity etc.), which causes the coating substance to spread evenly over the entire surface by means of centrifugal force effects.

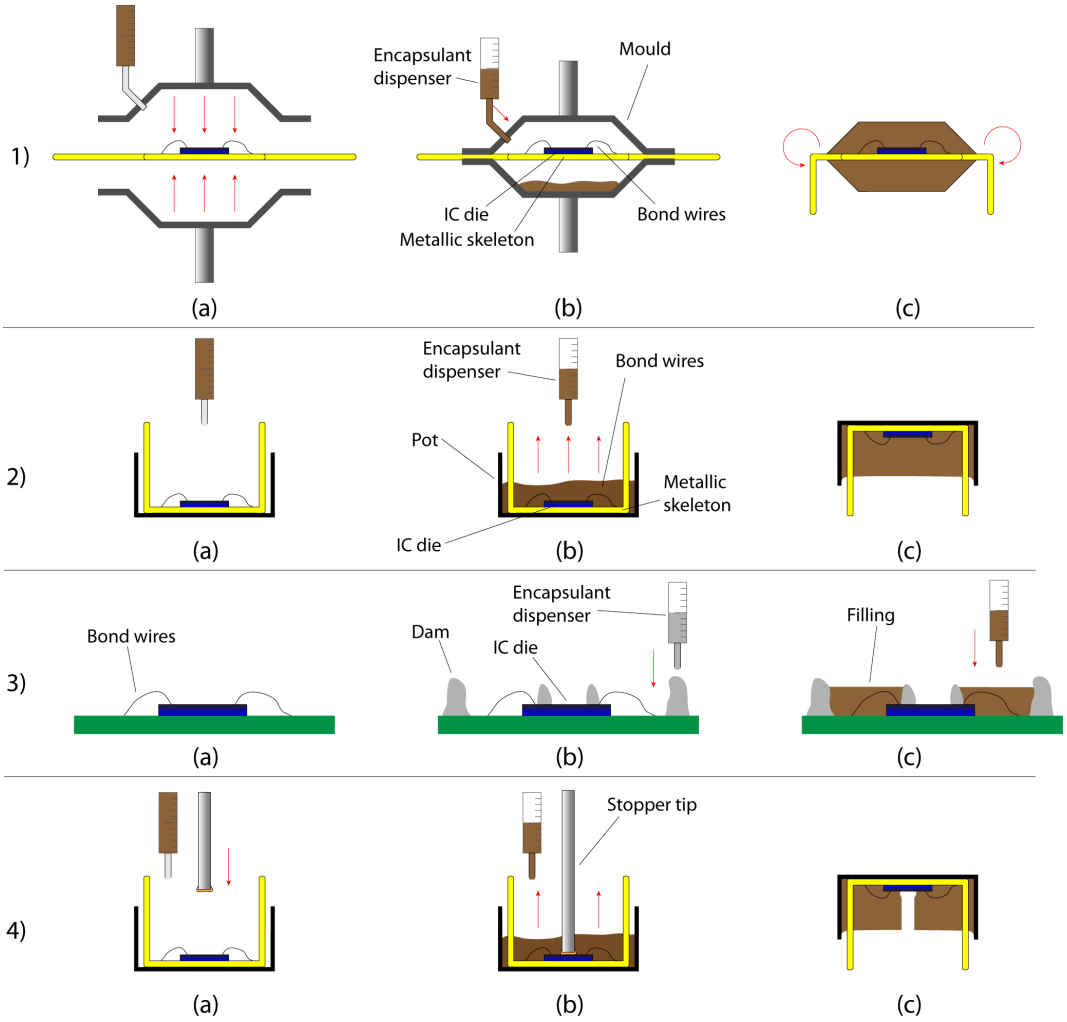

Figure 2.1: The mainstream types of encapsulation. Shown are: 1) Transfer moulding. 2) Potting. 3) Glob-top dam & fill version. 4) Potting with stopper tip (not actual mainstream method but possible modification to allow for selective encapsulation). Transfer moulding: (a) Shutting the mould over the wire-bonded die. In this depiction the die is sitting atop a metallic structure that also forms the external pins of the package. (b) Injecting encapsulant into the mould. (c) Post-processing of the cooled ensemble (e.g. bending the pins into the correct position etc.). Potting: (a) Mounted and bonded die sitting within an empty 'pot'. (b) Injecting epoxy encapsulant into the pot. (c) Filled pot ready for post-processing (e.g. painting the component specifications on the outside fo the pot etc.). Glob-top dam & fill: (a) Die mounted on top of a conventional PCB and wire-bonded. (b) Depositing the dam. (c) Depositing the filling in order to create the finished product. Potting with stopper tip: (a) Lowering the stopper tip. (b) Releasing the encapsulant material. (c) Removing the stopper tip to reveal the finished product.

in this paper exceed the 1 square cm mark in area. Miniaturising this technique to dies a few mm across would appear to be a rather challenging task. Potential issues might arise mainly due to wasted chip area since the gasket has finite thickness and has to sit well inside the pad-ring or risk damaging the bond wires.

2.1.2.4 Sacrificial ink approach

Yet another method that attempts to streamline the encapsulation of microfluidic dies is presented in this paper [13] and is based on the concept of 'sacrificial ink'. In this more intricate process the die is first wirebonded normally and then a highly viscous encapsulant (semi-cured resin) is poured over the bond wires. The high viscosity of the encapsulant in combination with surface tension are relied upon to force the semi-cured resin to flow along the bond wires whilst stopping short of spilling over the central part of the die. Once that is completed, the die is profiled by a laser imager and a height map is produced. Based on that map an ink-dispensing micronozzle ejects ink atop the die thus forming a string. This ink runs between a microfluidic inlet and an outlet forming a sort of continous tube between them. Finally, the resulting set-up is 'dammed' and sheathed in a low viscosity encapsulant and the cured product is heated to a fairly low temperature by semiconductor standards (the authors mention approx. $75C^o$) and the now melted ink tube is removed to reveal a microfluidic channel. Although the authors mention the flexibility of the method in so much that it can create complicated and fully custom shapes, the amount of effort and equipment required to achieve this is considerable.

2.1.3 Summary

Unfortunately encapsulation-based packaging methods do not come without an array of problems.

To begin with, there are always issues with encapsulant permeability to water. This book on microelectronic packaging [7] highlights the problem (Section 15.9) by illustrating the water permeabilities (a.k.a hydraulic conductances) of materials along the same axis. The image is reproduced further below for the convenience of the reader (Figure 2.2)[3].

Note: Hydraulic areal conductance here represents mass flux density (units of g/sm^2, the equivalent of current density in electronics) over pressure difference (units of Pa, the equivalent of voltage difference in electronics); a quantity that is a function of material thickness (plotted on the y-axis) and the material's inherent hydraulic conductivity. Important note: the terms 'conductivity' and 'conductance' are considered with their electronics meaning and transferred to the hydraulic equivalent. For example hydraulic conductivity in the electronics sense of the term 'conductivity' links the equivalent of current, mass flux (g/s) to the equivalent of voltage, pressure difference ΔP and device geometry (area and length): $g_e = \frac{I}{V} \cdot \frac{m}{m^2} = \frac{g/s}{Pa} \cdot \frac{m}{m^2}$ with S.I. units of s, where g_e denotes conductivity in the electronics sense of the word. In the fluidics sense hydraulic conductivity links pressure difference to fluid velocity through the medium via the fluid properties (the nature of the 'fluidic current' particles - equivalent to electrons in electronics) and the material length Δx as expressed with the help of Darcy's law: $K = \rho g \frac{v}{\Delta P/\Delta x}$ in units of m/s, where ρ is the desnsity of the fluid, g the acceleration of gravity and v represents fluid velocity.

[3] In the original the y-axis is unlabelled and shows an unknown quantiy and the units of permeability are of incorrect dimensions ($\frac{g}{cm \cdot Torr}$ instead of m^2).

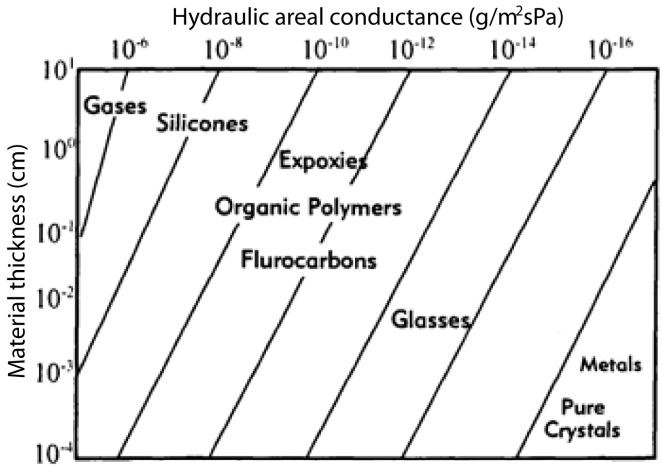

Figure 2.2: Plot of hydraulic areal conductance of various materials vs material thickness. Adapted from [7].

Details aside, the key concept hinted at in Figure 2.2 is that with time water will penetrate the various epoxy resins and Silicon compounds used for encapsulation and settle in areas where bond-wires are present. This problem affects all encapsulation-based packaging approaches and will eventually cause serious failures. The die itself will often get contaminated, but generally speaking the bond-wires shorting with each other is the more immediate concern, as evidenced by ISFET carrier chips that operate while routinely in contact with aqueous solutions.

Another severe problem that dies (particularly large ones) are vulnerable to is 'chip warping' whereby the difference in thermal expansion coefficients between die and encapsulant creates stress on the die as the freshly encapsulated die cools down from curing temperature to room or nominal operation temperature. This is also a serious issue that affects all encapsulants except those that cure at temperatures close to room or nominal operational temperatures. Once again the gravity of the issue is evidenced by research undertaken in the field. This extremely specialised paper is a very good sample of such research [14].

Yet another difficulty arising from the use of encapsulants is wire-sweep. When encapsulant is inserted incorrectly or too forcefully, bond wires may be displaced by the incoming torrent of molten material and in extreme cases end up touching each other. This can affect any method where the encapsulating material is injected under high pressure and in bulk. The issue of wire-sweep within the context of transfer moulding is discussed in this paper: [15].

Finally, a few more problems that various methods ail from have to do with planarisation with an eye to attaching a microfluidic channel atop the complex (glob-top experiences this problem very severely), the use of bulky equipment (transfer moulding in particular suffers from this, but it's by no means the only technique that faces that issue) and the execution of awkward 1-time operations (photoresistive methods in particular are badly affected). [1, 7] go over some of the sources of problems whilst this

Table 2.1: Summary of basic packaging techniques and reasons to use and avoid each method. Most reasons are mentioned in [1], others are based on observations from the literature.

Technique	**Reasons to use**	**Reasons to avoid**
Transfer moulding	• Fast • Repeatable • Can be performed selectively	• Causes chip warping • Requires bulky equipment • May cause wire sweep
Potting	• Simplicity • Fast • Uses pre-fabricated case so no mould is required	• Causes chip warping* • Risk of air-gaps/bubbles†
Glob-top	• Easily automated • Good selectivity	• No mould or case present in advance s repeatability suffers • May cause warping*

*When used in conjunction with hot epoxy
†Unless done under high pressure (which would turn it effectively into transfer moulding)

paper [16] showcases a study of 9 encapsulant materials for important properties.

2.1.4 Other work

More research discussing the problem of encapsulation of biosensors that need part of the die exposed to the elements can be found in [17]. Encapsulation of an ISFET-carrying chip with the aid of photoresist patterning is discussed in [18]. A short treatise on microsensor packaging with an example taken from thermal imagers can be found in [19]. The full paper detailing the construction of the IR thermosensor from [19] with additional information on chemical, thermal and stress sensors and their encapsulation is [20]. The same IR thermosensor can also be found in a publication with more detail given to the packaging aspect [21]. Work combining a conventional dam & fill approach and biocompatible encapsuation for a nerve stimulator can be found in [22]. The practical implementation of dam & fill techniques in a high-throughput, low-cost industrial environment is studied in this commercial reference: [23].

Furthermore, a large number of research publications consider the encapsulation and packaging of their experimental dies as important enough to mention in the final publication. Often researchers come up with very creative ways of packaging their dies. A couple of examples can be found here: [24, 25].

A summary of encapsulation methods can be found in Tables 2.1 and 2.2.

2.2 Inductive power & data telemetry

In various applications integrated circuits need to communicate data, power or both with the environment outside them completely wirelessly. This often imposes severe restrictions with regards to the use of off-chip components. Specifically off-chip components are either completely prohibited, or should reside within a very confined space, possibly sharing package with their corresponding die. Implantable applications are particularly good examples of systems that are subject to such restrictions. For that reason

Table 2.2: Summary of advanced packaging techniques (continued from Table 2.1).

Technique	**Reasons to use**	**Reasons to avoid**
Photolithography	• Good scalability (wafer scale) • Good selectivity • Good repeatability	• Requires specialised equipment • Hard to process individual dies
Gasketing	• Simplicity • Low cost • Easily reversible	• Requires precise alignment • Not suited for small dies
Sacrificial ink	• Good selectivity • Integration with microfluidics	• Complex process
Capillarity-based	• Good biocompatibility • Good selectivity	• Complex process

research into integrated circuit telemetry has been intense. Over time this research has led scientists down a couple of main paths: inductive coupling and optical coupling. In this section we shall examine the former. Note: capcitive coupling has also been considered and a small subsection will be devoted to it within this section.

Inductive coupling has been the most obvious way to create telemetric systems. The theory revolves around the exploitation of electromagnetic coupling between a pair of coils. Specifically, if an AC voltage is applied across the terminals of an inductor, a magnetic field will be created through the core of the said inductor and also 'spill out' around it in a pattern that can be calculated from the geometry of the set-up and the application of the Biot-Savart law[4]. This magnetic field can then be captured by a nearby receiver inductor and converted into an electromotive force (EMF), according to Faraday's law of induction (see Figure 2.3 for illustration). Within the context of practical telemetric systems the nomenclature states that a 'primary coil' sends power and a 'secondary coil' receives it. The sizing, alignment and configuration of the coils and the frequency of operation are typical engineerable parameters of this system the aim being to maximise the 'coupling coefficient' between transmitter and receiver coils towards the maximum value of 1 (100%) in such applications[5]. Of course the coils also need to be connected to appropriate driving or power management circuits to ensure performance in a practical set-up.

As such, this research can be classified into the following categories:

- Key milestones: Research that has helped start off the field of inductively coupled microelectronic systems or brought about important advances in the field.
- Technological papers and circuitry: On the technology side intelligent ways of exploiting CMOS

[4]Biot-Savart law: $\vec{B} = \frac{\mu_0}{4\pi} \int_C \frac{I d\vec{l} \times \vec{r}}{|\vec{r}^3|}$ where $\vec{B}$ is the magnetic field, μ_0 the magnetic permeability of vacuum (and by approximation also of air), C is the electrical current path through the circuit element that generates the bulk of the magnetic field, I is the current through the aforementioned circuit element, $\vec{l}$ is the directional, differenitial current path element (magnitude: length of element, direction: conventional current flow) and $\vec{r}$ is the displacement vector linking the piece of wire corresponding to $\vec{l}$ to the point in space where the magnetic field is being computed. This equation allows the calculation of the magnetic field flux vector at any point in space given a specified current flow path and current magnitude.

[5]Equivalent to trying to maximise 'mutual inductance', a closely related concept. Whereas the coupling coefficient varies between 0 and 1, mutual inductance is in principle unbounded.

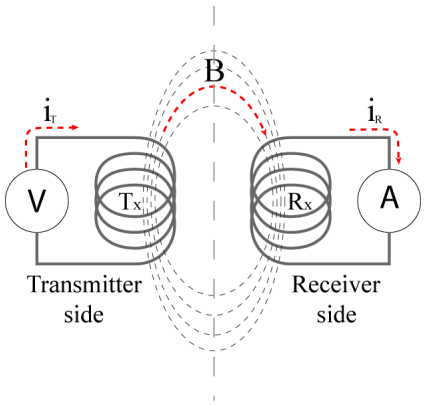

Figure 2.3: Concept of inductive coupling. In a typical system, a coil driver applies a voltage across the inductor, typically within the context of a more complicated circuit. A current, i_T starts flowing in the transmitter coil (T_x) and a magnetic field B is generated as a result. This magnetic field is intercepted by the receiver coil (R_x) where it creates an induced current i_R. This current can then be detected by various means, in this example an Ammeter. The induced current can be used either as a signal or harvested for power.

technology and various post-processing steps is used to improve the efficiency of the inductively coupled systems. Layout techniques are a good example of such engineering. On the circuitry side, research dealing with circuits that are intended to aid inductively coupled systems recover power, receive signals or emit signals more efficiently are included. For power recovery conversion efficiency is important whereas for signals link speed and SNR also play important roles. This section also includes papers operating more in the domain of 'analogue software', for example on the topic of efficient communication protocols. The two subcategories are incorporated into a single entity because very often papers tackling layout also introduce new circuitry and vice versa. This is particularly true for work emanating from Keio University.

- Application-specific implementations: A wide variety of applications that rely on inductive telemetry are presented in the publications residing in this section.
- Theoretical work: Specialised theoretical work relating to the application of inductive coupling in practice. Here we concentrate on theory concerning applications within the framework of integrated circuits.
- Review papers: Publications providing useful overviews of the field.

2.2.1 Key milestones

In this subsection of the thesis a small number of papers is examined. These papers were considered by the author of this thesis as key milestones in the development of integrated electronic systems that use inductive coupling for the transfer of power, data or both (within the context of this PhD). The literature in this domain is extremely extensive, so only part of it has been reviewed with particular stress on single-IC systems using as few external components as possible (typically the antenna).

A good start would be research published in 1997 [26] that demonstrated a very simple system capable of receiving power wirelessly, modulating a signal that would arrive from a sensor and then transmitting

the resulting data to an external receiver. The researchers quote a data rate of 62.5 kHz for a system that operates in single-digit mW domain in terms of power. This infant system features a pair of coils whereby one member of the pair handles power transfer tasks whilst the other handles data. All coils were manufactured by post-processing applied on the test dies. The set of references quoted in the publication consists of a paper detailing how inductive coils can be added to the back-end of an integrated circuit via post-processing [27] and another one exhibiting design methodologies for micro-coils aimed at wireless power transmission in microsystems [28]. This publication, however, seems to mark the first time when a fully functional and 'complete' system has been reported on the principles of inductive coupling. By 'complete' it is meant that the system is capable of achieving power transfer, data read-in and data read-out based solely on the phenomenon of inductive coupling.

A very early full system application positioned on a full-inductive system took the form of a microstimulator chip [29]. The coil is not integrated, an external coil being used instead, however the matter of fact is that a full microstimulator driver is implemented in a system that has no galvanic contact with the original data source and all this runs on a power budget of 14.8 mW and a data rate of 8.3 kbps according to the authors. This is already not unreasonably high given the power that has been transferred inductively in the previous reference [26].

By 2004 integrated inductive systems had begun using the FSK (Frequency Shift Keying) communication protocol [30]. This seems to be a staple of modern day inductive communication systems, and the cited reference is an example of a system that uses it which is why this paper resides in the category of key references[6]. The concept revolves around the idea that since a carrier wave is necessary to transmit power to the system, rather than attempting to avoid that frequency when dealing with data and going through all the effort of building separate coils for power and for data, it is much better to simply modulate the carrier frequency in order to transmit data; a system much akin to FM modulation in the radio. Interestingly, this work shows dramatic improvement over previous attempts [31] by achieving 3 Mbps although it can also operate at low power mode by consuming less than 0.5 mW if ran at 200 kbps, according to the specifications summary table in the paper.

At this point, another scientific group enters the scene. Keio university's Kuroda-sensei and the rest of the group have been very dynamic players in the field of inductively coupled chips with a lot of work done on wireless buses. The true power of inductive coupling was revealed when it was used for the implementation of a wireless data bus [32]. A test system was manufactured and presented along with a meticulous treatise on the design of the set-up, both at circuit level and at physical level, and on cross-talk considerations. Moreover, a new signalling scheme called NRZ (Non-Return to Zero) was introduced. The same group followed it up with dedicated papers later on. Note: [32] seems to be a more consolidated version of this paper [33], but still illustrates the concept of the wireless bus very well.

From the world of bio-electronics inductive coupling was used in order to recover an EEG signal from an implant that required no wires to the outside world for its operation [34]. This represented an early attempt to utilise inductive technology directly *in-vivo*. The added difficulty involved in the task of capturing and processing with accuracy signals that are in the micro- to mili-Volt range should not be underestimated.

The issue of cross-talk in inductive systems was bound to generate research sooner rather than later

[6]Note: although there is some movement towards UWB protocols (from verbal communication with a colleague -Song, Luan- from the field at Imperial College) FSK has played an important role in the field of inductively coupled systems.

given the well-known property of EM fields to 'spill over' into the area around them. When an inductive coil is generating an EM field, the flux density of the field drops quickly whilst its geometric extent spreads. This phenomenon was studied quantitatively within the context of arrays of channels laid-out side-by-side [35]. Attempts to implement techniques that would minimise the disruption at minimal cost of layout area have also been developed with some already appearing in the aforementioned [35].

By 2003 the idea of designing a telemetry chip for use in tandem with implantable microsystems was being examined [36]. Apart from a theoretical treatise of antenna operation and descriptions of the design and construction of the prototype system a test was run where the quality of the connection was compared between a set-up where telemetric transmission had to occur through an air-gap and a variation of the set-up, where the space between the primary and secondary coils had been filled with tap water to at least partially emulate the effect of tissue. The verdict was that fortunately RF radiation is not heavily attenuated when passing through water in the sort of quantities that are expected when dealing with bio-electronics.

Another key advance in the field was constituted by the integration of an RFID system into commercially available $0.18\,\mu$ technology [37]. The interesting point here is that the entire system, including the antenna coil is integrated into standard CMOS Silicon. Previous work tends to use post-processing to add a vast antenna made of very thick metal on top of the finalised die. Although within the context of the reference above the system was not used in 'operational conditions', that is to say demonstrated to work as a system, its vital statistics were examined bit by bit until the conclusion was reached that the system was indeed (potentially) fully operational. Thus a precedent was set for using inductive coupling for power and data telemetry in very high-throughput standard CMOS with none of the complications of post-processing. Whether this has resulted in sacrifice of performance or not is unclear from the paper itself.

On a similar note a fully integrated system with inductive power delivery has been presented where this system now includes a power management unit consisting of a rectifier, a voltage reference and a linear regulator with the capability of providing a 'decent' (i.e. stabilised by regulator and sufficient to drive a wide range of load circuits) power supply to a load circuit [38]. Thus, this research extended the idea from the previous reference [37] to the domain of general electronics: No longer does the system need to be an RFID tag; now it can include generic electronics. The power output of the system is quoted by the authors to be 8.5 mW (regulated DC output) provided that the input power is of the order of 40 mW. This is more than enough for a huge array of applications, particularly given the trend of shifting more and more functionality into micro- and nano-power systems (so well below mW territory). In the paper, a practical example of an application is demonstrated by performing frequency variation analysis on a 57-stage ring oscillator.

The last piece of the puzzle was added when a fully integrated system with both data and power telemetry capability (including suitable circuits) was implemented [39]. The result was sadly not a physical die, yet the simulated results proved the concept [7]. It is noteworthy that both the power supply and the data communication modules presented in the publication are generic enough to service a large variety of systems that may potentially reside on the same die.

Finally, from a commercial point of view, [40] shows an example of single-chip inductively coupled

[7] The system was eventually realised physically but the results from it have not been published. From verbal communication with the author's supervisor -Dr. Timothy Constandinou-.

system research that has made it into commercial application, even though it does use an external antenna.

2.2.2 Technology of inductive power/data telemetry and associated circuitry

In this subsection, we shall be dealing with work that has explored the electronics manufacturing technology and applied this to inductive systems. Manufacturing procedures and layout techniques are included in this section as is associated circuitry and signal processing designed to allow inductive systems to operate more efficiently.

2.2.2.1 Manufacturing procedures and layout for inductive systems

Casting an eye to the origins of the whole field we find research dating back in 1997 where a couple of materials (NiFe and Cu) and a few manufacturing topologies are examined for efficiency [41]. This directly follows a couple of papers that have already been mentioned in the key category [42, 43]. Soon afterwards, people started thinking about using this new technology in transformer applications where the secondary coil is to be found on-chip [44]. This work [45] details the manufacturing procedure called 'inlaid electroplating' used to create inductive coils at the post-processing stage and tackles theory related to the efficiency of power transmission. Since those days CMOS technology has evolved further and now a range of manufacturers offer analogue (or RF) top metal options. These top metals tend to be many times thicker than their more deeply buried counterparts and are intended for use as inductors (and often convenient for use as high power supplies), thus replacing the post-processed antennae of the past.

In terms of layout there has been research, of which a couple of aspects are examined here. A simple, yet effective multi-helix inductor topology whereby more than 1 metal layer is used in order to create a multi-turn inductor figures as a possible layout technique [46] for compact inductor design. An architecture whereby a power supply inductor encases a 'clover' of data transmission inductors for the purpose of eliminating interference from the power coil has also been proposed [47]. This 'nested clover' configuration requires a closer examination: The data inductors consist of a pair of turns of opposite chirality. Therfore when a 'common mode' signal passes through the power transmission coil it generates power through the power coil but creates more or less equal and opposite eddies in the pair of turns of the data coil. That way localised signals have to only affect 1 side of the data coil (or affect both sides of the coil, but with opposite polarity) can force a strong signal through the data coil.

2.2.2.2 Signal processing and circuitry for inductive systems

At the same time efforts have been made to reduce the power necessary in order to run the telemetry with innovations such as NRZ (Non Return to Zero) signalling and a scheme implemented in order to control the transmitter power so that data is transmitted with sufficient reliability, yet at the same time without excessive power dissipation [48] (for details see in paper). NRZ signalling features in this follow-up to [32], alongside other techniques. The circuitry needed to implement NRZ signalling is described in this paper [49]. Over the years, progress has been made on the design of a system with many parallel inductive channels featuring a dynamic power controller in an effort to further limit the power dissipation of the link [50]. Later on, further attempts to save power led researchers to the implementation of a

daisy-chained architecture [51] between data transmission channels that would allow the die to make extensive use of charge recycling. As has been reported in [51] this effort was not trivial: It required careful consideration of design and layout of the inductors and assorted circuits, for charge recycling comes at the price of signal attenuation between stages and timing skews. The authoring group reported a sizeable reduction in energy per bit vs. a comparable system made in the same technological node (65fJ/bit vs. 140fJ/bit for the older work). This figure was subsequently overatken by work presented in [52] where the authors have implemented a biphasic pulse circuit that can create a signal of given amplitude with 1/2 of the power supply voltage that is normally required to drive monophasic signals. The reported figures are 20fJ/bit for data and 135fJ/cycle for clock efficiency.

There have also been efforts to study and eventually improve the basic power management systems that inductively powered systems require. This way the less power hungry systems detailed in the previous paragraph would gain access to power from more efficient PMUs (Power Management Units). Any PMU will typically feature some type of rectifier. This research paper [53] tackles the issue of integrated rectifiers built from different types of transistors (BJT, CMOS, or a mixture of both) into a fair amount of depth. Rectifier topologies, both circuit-wise and geometrically are also treated as parameters that affect the efficiency of the overall system. Onwards, a full PMU system is presented in [54]. A slightly more advanced system is reported in [55], additionally describing the design of a full front-end for RFID (Radio Frequency IDentification) systems. Meanwhile the problem that arises because inductive power delivery systems attempt to create very high voltages at the receiver coil was noticed. Circuit designers are often exhorted to use inefficient methods such as forced clipping in order to protect their circuits from the said high voltage. To counter that the implementation of a PMU in high voltage (HV) technology or dedicated HV devices within a technology before generating a stable low voltage power supply for the rest of the chip was proposed [56]. Note: This research could equally well reside in the power management circuitry section of this literature review, however since the implementation of these systems is specifically for use in inductive systems, it was placed here.

From around 2004 the potential of inductive coupling in microelectronics begins to become reality as literature starts appearing where the focus of the research is the improvement of signal transmission and not just power. These works [57, 58] are follow-ups on [32] where the focus is on the mathematical modelling and optimisation of the layout for the actual inductors used in the system, on the circuits necessary to implement NRZ signalling, a majority vote signal receiver etc. The data rate quoted for this work is a maximum of 1.25 Gbps/channel. Since that work there has been much literature dealing with ever increasing data rates. Notably many of the references also include the circuitry used to achieve the said data rate and very often measured results relating to issues such as misalignment tolerance, speed performance vs. BER (Bit Error Rate) etc. This literature is summarised in Table 2.3. Crucially, the table tells us that if we are willing to pay the energetic price we can get more and more speed and reliability out of our systems as time progresses. On the other hand we can always opt to sink power dissipation while maintaining an acceptable reliability and speed. Other relevant work includes an improved demodulator for BPSK-encoded (Binary Phase Shift Keying) signals [59], a method of signal transmission based on 'bursts' [60] (i.e. sending a signal and a signal-specific clocking sequence together in an asynchronous batch as opposed to employing a vast array of parallel, synchronous channels) etc.

Table 2.3: Indicative overview of the literature on inductive coupling detailing the increase of bitrates and decrease of bit error rates (BER) over time.

Ref.	Data-rate (Gbps/ch)	BER/Channel (max)	Year	Special innovations
[33]	1.2	-	2004	-
[32]	1.2	-	2004	-
[30]	0.0025	1.00E-4	2004	FSK protocol presented
[49]	1	-	2005	-
[61]	2.8	-	2005	-
[62]	≈ 1	1.00E-16	2006	Massive parallelism
[50]	1	1.00E-12	2006	-
[51]	1	1.00E-12	2008	65 fJ/bit
[60]	11	1.00E-14	2008	Burst signal transmission
[63]	0.15	1.00E-12	2009	Memory application Relatively good range
[64]	2	1.00E-12	2010	Stacking topologies discussed
[65]	8	1.00E-16	2010	Massive parallelism
[52]	1.1	1.00E-12	2010	Biphasic signalling. Lower power
[66]	12	1.00E-13	2011	Large single channel area
[67]	0.85	-	2012	Focuses on low power
[47]	6	1.00E-14	2012	Nested clover technique

Note: figures are merely indicative since both data rate and the BER vary with power and distance for example. For that reason the data rate/mW is often quoted as a metric. Nevertheless, this table indicates that if we are willing to pay the energetic price for it we can get more and more speed and reliability out of our systems as time progresses.

2.2.2.3 Inductive systems in a biomedical setting

An interesting sub-category of research within the field of inductive coupling is that of biomedical electronics. Because of the extra constraints and rather unique operating environment, research aimed at this particular domain often brings new perspective to the same issues that are encountered all over the field.

On the subject of power limitation an interesting perspective can be seen in [68] where the biomedical nature of the application imposes additional restrictions on power dissipation. No longer is the objective to send as much power as quickly as possible, but now the idea is to keep power dissipated by the implanted device within biologically safe limits. To that end the paper proposes a scheme whereby the power emitter and the implant communicate via a feedback loop. The intent is that once the implant detects that it is handling more power than is biologically safe, it will instruct the external power emitter to temporariliy shut down. This method can ensure that the power transmitted to the implant remains on average within the rigid specifications imposed by safety constraints.

Other work related to bio-implantable electronic reports a, general-purpose telemetry IC that is studied under conditions that would match the operational environment of an implant [69]. Notably an effort is made to measure power transmission efficiency at distances that are more appropriate to implantable devices as opposed to stacked dies in communications systems, and signal transmission efficiency through a 'tissue simulant' (in this case gel slices of varying thicknesses) as opposed to simply an air-gap. Once again the verdict is that thin-enough tissue should not disrupt RF transmission outside

the boundaries of the acceptable.

Nevertheless, research published in 2011 indicated that if the implant were to be placed under the skull power efficiency would be severely compromised [70]. This observation is primarily based on case studies and simulations, however it does provide some insightful results that affect bio-implantable design. Notably the case studies make an effort to show how the presence of muscle and skull would affect the telemetry with some analysis on inductor geometry as well.

2.2.2.4 Other work

Finally a group of rather more 'exotic' work dealing with the technology and circuitry of inductive coupling includes, for example, stacking techniques for multi-die towers have been examined [64], a massively scaled up a telemetry system resulting in 1024 channels with an aggregate bitrate of 8TB/s has been designed [65] and a single, large inductive coil that is time-shared between the signals has been implemented [66]. The objective here is to allow systems to retain their data transmission rates while improving the range of communication (from 10s or 100s of microns to few mm). In exchange the internal speed of the system needs to be increased in order to allow this time-sharing scheme to operate at a good aggregate bit-rate. A final example is this publication [71] where an effort is made to start providing EDA support for automated inductor layout.

2.2.3 Inductive coupling in practical applications

This subsection presents work that has demonstrated inductive telemetry within the setting of a practical application. The purpose is to put the entire field into perspective from the most practical point of view.

An early application of inductive telemetry where an -admittedly- external coil is used to handle communications with the outside world can be seen in [31]. The system was designed to be compatible with real microstimulator probes with an eye towards neural applications. The authors quote data rates in the region of 100 kbps and power dissipation in the region of 10mW (Note: the technological node is $3\,\mu m$ BiCMOS).

A number of years later a system that could theoretically be used to stimulate the visual cortex in order to restore part of the visual function to profoundly blind people was introduced [72]. The paper presents a variety of innovative concepts, however the key issue with regards to the context of this thesis was the use of inductive telemetry. The authors report 0.88 mW power dissipation for data rates in the Mbps region (0.2-1.5 according to the summary table). The implant is designed for manufacturing in 0.18 micron technology.

Other applications include a simple, integrated humidity sensor [73], a completely sealed memory chip stack [63] and a non-contact wafer testing platform [74].

In biomedical systems inductive coupling has been used for cochlear implants as early as 1989, if not even before [75] (although by unrestricted use of external components -diodes, resistors, capacitors, antenna, transformer-, so somewhat outside the focus of this literature review). Similarly inductive coupling used for retinal implants is not new (for a 1999 example [76]).

2.2.4 Theoretical work

This section presents a small number of samples pertaining to relevant theoretical work that has been reported over time. These specimens of research include detailed mathematical analysis and focus on theoretical calculation, simulation and some times measurements in order to verify the simulated or calculated results. The purpose of this section is more to illustrate the vast range of topics that research has been conducted upon within the wider context of inductive systems.

To start with, a publication proposing the implementation of an Unscented Kalman Filter (UKF) in order to improve the sensitivity of a passive telemetry sensor system (PTSS) has been identified [77]. The publication is heavily mathematical and includes simulated data. The UKF first models and then cancels out imperfections in the signal brought about by non-linearities introduced by the telemetric link itself, and is thus directly relevant to the inductive coupling topic.

Other theoretical work includes [78] where a number of on-chip inductor topologies were studied mathematically, numerically analysed and then compared to measurement results. The cornerstone of this work seems to consist of what the group calls 'differential multistacked' structures that basically consist of spiral inductors spanning more metal layers. This research is conceptually paired with similar research presented in [79], where analysis follows the same line of thought, but different topologies are considered.

Finally, a study on misalignment from a mathematical point of view has been identified [80]. The publication can help engineers visualise how misalignment in various directions (square coils are considered so the geometry is not isotropic along the plane of the inductors) and at various distances is expected to alter and corrupt the communication channel.

2.2.5 Review papers

This subsection presents a handful of papers reviewing the field of inductive telemetry (the papers are not necessarily 'review papers' from a technical point of view but carry such value nonetheless).

To begin with a short, early (2005) review paper presents some basics of RF communications and lists out a number of practical considerations when creating a real system [81]. Specifically some equations describing the electric and magnetic fields created by an inductor are shown, some comments on near-field coupling vs. far-field coupling are shared with the readers and the details of the manufacturing of a simple Colpitts oscillator-based system are presented. Finally, the test set-up is presented and results for telemetry through air and through water are reported.

Next, this work from Keio university [62, 82] gives a very brief overview of TSV, micro-bump, capacitive coupling and inductive coupling as methods that can be used to pass signals between different dies. Then, a sample inductive system is presented (work done by the same Keio Univ. group) as a solution that can withstand competition from other approaches in a variety of applications.

A short and succinct review paper dealing with generic chip-to-chip communications, where the authors considered briefly the perspectives of wired, optical, inductive and capacitive coupling as possible solutions to the data-rate conundrum (the complexity of the connectivity pattern tends to grow much faster than the complexity of the underlying system much like the number of edges in a complete graph

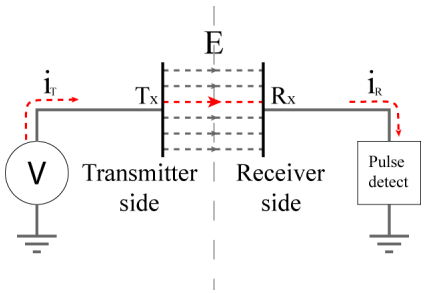

Figure 2.4: Concept of capacitive coupling. In a typical system a capacitor driver applies a voltage with respect to ground to the transmitter-side plate of the communications capacitor (T_x), typically within the context of a more elaborate circuit. Charge is transferred to (or from) the transmitter plate, giving rise to a transient current i_T. This creates an electric field across the capacitor plates (E) which causes charge to shift away from (or towards) the receiver plate (R_x). The current that arises from this shift , i_R can be used to create transient voltage fluctuations which can then be detected by a votlage pulse detector (for example a dynamic-logic-style circuit). i_R can be used either as a signal, or as a power source.

grows much faster than the number of vertices) has been identified [83].

'Enabling technologies' for the field of what the authors call 'proximity communication' are reviewed in [84]. Particular care is shown towards packaging that can prevent misalignment of the dies that are meant to be communicating this way although methods for detecting misalignment and subsequently correcting it (which can be done with a few purely electronic methods described in the paper).

Finally yet another publication from Keio Univ. discusses the basics of inductive communications providing interesting material such as images that help visualise electromagnetic (EM) fields in near and far field conditions, graphs plotting the connection strength vs. distance between coils, circuitry used in inductive systems, an image detailing the downscaling equations in inductive systems in comparison to Moore downscaling, an example working system etc. The paper concentrates on pulse-based communications but does form a good review paper in the process [85].

2.2.6 Capacitive coupling

Capacitive coupling is a subset within the topic of inter-chip communications that has received less attention than inductive coupling. The concept revolves around having each plate of a capacitor on different super-structures (die to die or die to board for example). That way every time the potential on 1 plate is brusquely changed the other plate will show some transient response. The impedance to which these capacitors are connected will play a role in determining the exact nature of this transient, but for the purposes of transmitting digital information a pair of distinct and reliably detectable transient events to represent a 0 and 1 may be all that is needed. To see an illustration of the principle of capacitive coupling consult Figure 2.4.

There has been some work done in the field, but typically problems with range limit its utility. In the review subsection of the inductive coupling section (2.2.5) some of the papers discuss capacitive coupling [62, 82], as was mentioned in the paragraphs describing them. The main problem cited typically is that it can only work for a couple of very well-aligned dies that are facing each other. Otherwise the linkage is simply too weak.

The concept of on-chip capacitive coupling has been around since at least the 90's as can be proven by [86]. The authors consider the possibilities of capacitive coupling as a good way to make any interconnect, not just chip-to-board, less energetically taxing and cheaper.

An early, publication where a capacitive and an inductive interconnect are tested and described within the same work can be found here [87]. The systems are simple and the innovation revolves around basic characterisation of inductive and capacitive links. Some more advanced, yet still by modern standards basic capacitive systems where the design, simulation and characterisation of an early capacitively coupled system can be found in [42]. A similar system which stresses the importance of alignment between the dies has also been studied [43].

An example of capacitive coupling in action is given in a 2006 paper by Culurciello and Andreou [88], where both power and data are transferred through capacitors. The system is implemented in 0.5 micron technology and features a form of power converter that can take in the raw capacitive signal and convert it to a useful power supply. The authors report up to 15 MHz measured data transmission rate, 100 MHz simulated data transmission rate and a capability to source at least 9 mA down a 3.3 V supply although a maximum limit was not specified.

Generic issues related to capacitive coupling have been discussed and some circuits that could be used in order to develop a generic data transfer interface for capacitively coupled dies have been proposed in the literature [89]. The assembly issue was also considered in the aforementioned publication and the capacitive coupling method was compared with that of inductive coupling; a comparison that does not generally favour capacitive coupling but in this case does leave some room for application, possibly. The main issue seems to be that even though capacitive coupling does seem able ot match the data costs seen in inductive systems (fJ/bit), it severely lags behind in terms of area (fJ/bit/μm^2).

Capacitive coupling finds an application in [90, 91] where a system was designed with an eye towards on-wafer testing. The paper comes complete with circuit diagrams and measured results from a fabricated die, reporting a 2 Gbps data rate and also voltage sensitivity of 25mV.

The final example here identifies that due to the limited range of capacitive coupling the signal emitted at the capacitive interface is less likely to disturb other circuitry residing nearby [92]. This work is targeted towards bio-implant applications and comes with basic circuitry to achieve the telemetry and measured results.

As such we can see that capacitive coupling has potential for use in certain applications although a significant amount of research needs to be undertaken before such technology becomes comparable in performance with inductive links and commercially viable, as some of the authors admit themselves (for example in [89]).

2.2.7 Summary

On the whole inductive coupling is clearly a very well-developed topic with a variety of applications and great potential for improvement as our capabilities to manufacture integrated circuits improve. It comes, just like everything else in the world of engineering, with its own set of strengths and weaknesses.

Some positive points that can be associated with inductive telemetry are as follows:

- Speed: with rates in the Gbps/channel inductive telemetry has clearly shown the potential to keep up (or at least nearly keep up) with the electronics side of the problem.

- Direction of communication: The way integrated circuits are manufactured make dies particularly amenable to the implementation of inductive links in the perpendicular direction with respect to the surface of the die. This figures on the list of positive aspects because in any direction parallel to the die physical interconnects can handle the communications. The perpendicular data path opens the way for inter-chip communications.

- Low power dissipation: This is debatable, but as we have seen some systems quote data communication costs in the range of tens to hundreds of fJ/b (e.g. [80, 60]). In terms of galvanic charge transfer this is equivalent to charging/discharging a capacitor of tens to hundreds of fF capacitance across a $\sim$1V power supply (charge from GND or discharge from VDD) as evidenced by setting $V = 1$ in the equation that describes switch capacitor energy transfer: $E = CV^2$ where E is the energy transferred, C is the capacitance and V the voltage difference between initial and final states. In practical terms data transfer cost of fJ/b is of the same order as toggling a few minimum size CMOS inverters across a few V power supply[8]. This must be seen within the context of data transmission outside the die when wired solutions are not necessarily available or indeed applicable.

- Good near field localisation: Once again a matter subject to different interpretations, however from the literature it transpires that 30 micron coils (square coils in this case) can communicate to similar size coils on a different die with relatively little cross-talk.

- Wireless nature: Inductive telemetry has been proven capable of handling both power and data transfer in a number of applications with great success. This renders galvanic connections between the die and the environment unnecessary (when indeed the general set-up allows for that solution), thus eliminating the need for a pad ring as well. This implies that the die can now be completely encased in passivation and that die area is potentially saved (subject to the constraint that the inductive coils do not consume more net area than a pad ring would[9]).

- Relative immunity to skin tissue effects: As seen in the literature review above attempting to transmit a signal through the skull results in heavy losses. Nevertheless in rather more accessible locations where skin is the only barrier (atop the sternum bone for example) such losses should not be a serious cause for concern. This is in contrast to optical coupling whereby scatter and absorption cause any beam to fade quickly. As for the skull, it would stop any optical beam, even those within the 'bio-telemetric window'.

On the other hand one must mention the following negative points associated with inductive telemetry:

[8] Of course the specific energy/bit values for any inverter will depend on the technology, power supply and specific input signal waveform employed. Our declaration is based on information from the standard digital cell library datasheet for IBM's 0.18 micron H18 technology under a supply of $1.8\,V$ and the assumption of a very fast input bit toggle time ($10\,ps$). We averaged the low-to-high and high-to-low transitions.

[9] Estimating the net area cost of the inductive structures is not a trivial matter to determine as it is subject to a variety of factors. For example, an integrated inductor takes space in the metallisation layers, so a question is whether it imposes interconnect constraints so that the die area has to grow to accommodate the said interconnects. Another question is whether these coils need to stay away from any particular part of the front end (sensitive devices or circuits for example). And so on...

- Sensitivity to misalignment: This is perhaps the biggest undoing of the technology. Dies that seek to communicate via inductive telemetry need to be aligned very accurately. This largely follows from the point that the near field is reasonably localised and can impose a real limit on the practical applicability of the technology although as we have seen this matter has not escaped the attention of engineers.
- Limited range of communication: At some point near-field turns into far-field. Unfortunately this happens very early for inductive systems. The size of the coil largely determines the range of communication. In a system with bidirectional telemetry, the smaller of the coils (if of unequal sizes) sets the bottleneck. As we have seen mm-range communication is rather uncommon in the field, all telemetry typically occurring over 10s-100s of microns instead for systems considered throughout our literature review. Going farther than that introduces power loss and cross-talk.
- Regulation of bandwidth: The RF spectrum is known to be cluttered and regulated by law. Complying to the regulatory framework may impose non-technical limitations to the technology. This will depend on the application and may be mitigated to some extent by the short range of devices in certain applications that do not require communication over more than a few cm (e.g. medical).
- Interference: When an EM field penetrates a die, it doesn't just create eddie currents in the coils that are designed to secure signal and power transmission, but rather in any line of metal that they affect. This may cause problems in circuitry although typically the frequency at which the inductive transmission occurs is avoided by the said internal circuitry.

2.3 CMOS photonics: photo-emission, electro-optical modulation and optical power recovery

Optical coupling (within the context of this thesis) is based on the principle of communicating data and power with a die via electromagnetic radiation in the UV-IR range through specialised structures on the die. As such, photosensitive and photo-modulating or photo-generating elements are required in order to meet all the functionality requirements of a telemetric system.

These issues will be studied in this section of the literature review, which is subdivided as follows:

- Power-capture structures: This reviews devices with the capability of capturing incoming light are studied. Their design, layout and even fabrication are considered. Of course when capturing optical power the purpose might be to provide actual electrical power to a PMU (Power Management System) or alternatively the incoming power may well be the actual signal. As such both devices designed for power and data recovery are considered together within this section.
- Light-emitting structures: Design, layout and fabrication of devices tasked with generating light in Silicon, a task known to be somewhat challenging in the photonics community due to the indirect bandgap of Silicon.
- Light modulation structures: Design, layout and fabrication of structures with the capability of modulating light that goes through them either in terms of amplitude, or phase or both. Also included is a bit of literature tackling the underlying theory regarding free-carrier absorption and free-carrier refraction as means of achieving light modulation.

2.3.1 Power capture structures

In this subsection, integrated power harvesting structures and topologies that are implementable in either simple or triple-well CMOS processes are studied. The literature in this respect is relatively small, but succinct.

To start with a Silicon-on-Sapphire process has been examined as a technology which would allow the creation of very long chains of series-connected photodiodes (100s in a chain) for the provision of high voltage to a solar-powered system [93]. The work described within the aforementioned reference is brief but forms an example that illustrates the point that obtaining solar power from an integrated cell is not something that the science community has only started contemplating recently (the publication dates from 1991). A similar system (series connected photodiodes) was experimentally demonstrated much later on (2008) , albeit with just 4 photodiodes in the chain [94].

Next comes a very important reference that has been cited very widely by the integrated solar power community: Arima and Ehara's 2006 work on solar battery configurations [95]. The paper includes a cross-sectional diagram of an integrated solar battery (adapted in Figure 2.5) whereby:

- The + terminal of the cell is formed by a volume of p-diffusion on N-well.
- The p-diff/N-well junction is optically active, that is to say light is allowed to fall on it and create photo-electricity.
- The N-well is then connected via an ohmic contact to the substrate. This is achieved by connecting an n-diffusion region residing on the N-well to a p-diffusion region residing on the substrate. The physics behind this relates to the fact that the depletion region created by these highly doped diffusion materials through the shorting wire is so narrow that tunneling effects dominate over thermal ones thereby negating the rectifying nature that characterises less heavily doped pn-junctions.
- The p-diff/n-diff junction is optically inactive, that is to say remains shaded by careful use of the back-end metallisation. Any photocurrent generated at that junction would counteract the efforts of the solar battery.
- Finally, an n-diffusion sits atop the substrate creating the last pn-junction in the system. The n-diffusion forms the negative terminal of the set-up.
- The n-diff/p-sub junction is also optically active and works together with the p-diff/N-well counterpart towards the generation of electricity.

The paper carries on to show the structure implemented in 0.35 micron CMOS, characterise the resulting solar cells and use them to power some simple test circuits that reside on the same die.

Work has also been undertaken in order to optimise the layout and configuration of photodiodes for power harvesters. To begin with, different geometries of p-diffusion/N-Well (p+/NW) junctions have already been examined and compared [96]. By arranging the geometry of the devices in such way as to obtain different perimeter-to-area ratios contributions of side-wall and areal junctions can be extracted. The relation of the obtained power with respect to how the parasitic N-Well/p-substrate (NW/p-sub) diode is connected has also been examined (same publication: [96]). On a similar note, useful data about

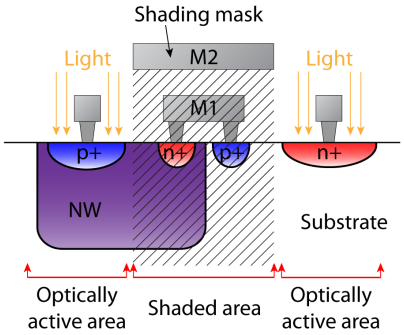

Figure 2.5: Solar battery structure proposed by Arima et al. [95]. Legend: M1: Metal layer 1. M2: Metal layer 2. p+: p-diffusion. n+: n-diffusion. NW: N-well.

different types of photodiodes (again p+/NW and NW/p-sub) including this time: spectral selectivity, responsivity, noise generation, SNR and fill factor is provided in [97], a publication in which the main focus revolves around imagers.

A new configuration of stacked photodiodes for the generation of a high voltage solar supply has been studied, where the main innovation revolves around a 'pyramidal' structure whereby the 'base' of the solar cell consists of a large number of unit photodiodes that feed into the 2nd stage in the series connection, which consists of far fewer unit photodiodes etc. In essence the series connected photodiodes are now of radically different sizes [98]. The objective was to compensate for losses via parasitic photodiodes. It is a well-known problem that when stacking photodiodes in processes allowing nested junctions the connectivity patterns of different junctions in the 'nest' or 'stack' can influence the measured phptocurrent output of the 'target' pn-junction (see Figure 2.6 for a fictive 6-well process). For example shorting an N-well on substrate junction (grounding both terminals) reduces the photocurrent measured from the p+/N-well junction. To make things worse, because every p+/N-well photodiode resides by definition in an N-Well there is no way of shading the N-well/p-sub parasitic photodiode without also shading the p+/N-well harvester one too.

The work of Arima and Ehara has been taken to the next level by recent work in which the original results were reproduced, but the possible pn pairings deriving from triple-well technology were added to the lists of results [99]. Therefore, photodiodes that are formed by use of the triple, p-type, well have now been characterised. The publication also includes a ring oscillator that would help illustrate the differences in performance stemming from different optical structures. A combined discussion about photdiodes along with some tests performed on creating multi-metal integrated capacitors with an eye towards more efficiently packed energy harvesting systems can be found in [100].

Apart from characterising and engineering photodiode structures themselves, a certain amount of work centred on combining them with other-purposes circuits has been done:

An interesting modification to CMOS circuits that allows energy harvesting out of those sections of pn-junctions that appear in MOSFETs but do not take active part in normal device operation (for example in an inverter the p-diffusion (connected to VDD)/n-well junction of the pMOS and the n-diffusion/p-substrate (connected to GND) of the nMOS are exploited [101]. A cross-sectional diagram of these devices within the technology can be found in the paper.

Imagers contain such pn-junction elements that perform the role of camera pixels. The concept of

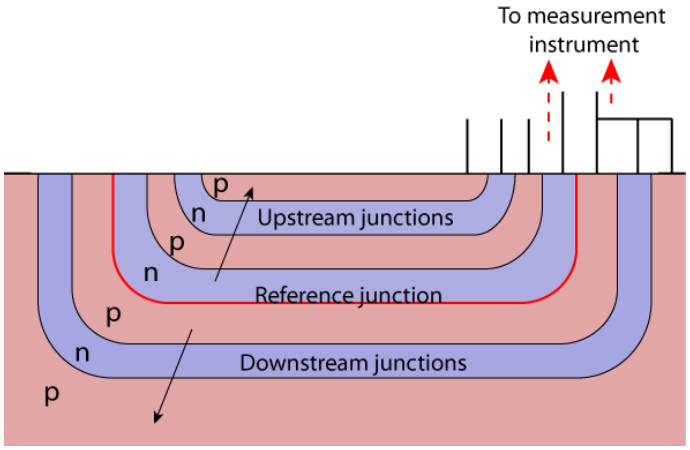

Figure 2.6: The challenge of stacked diodes in a multiple-well CMOS compatible process (in this generic example a fictive 6-well process). When measuring photocurrent generated across a target pn-junction within a junction stack the connectivity pattern of other junctions in the stack can influence results. For more information on how shorting various junctions affects the I-V characteristics of the reference junction see [99]. The figure shows an example of a 'junction cake' with the target junction highlighted in red, 'upstream' junctions left floating and 'downstream' junctions shorted to the substrate (GND).

using them as pixels whilst reading data out of them and using them as power recovery structures when idle has also been examined, leading to the creation of a self-powered imager demonstrator chip [102]. The test chip was implemented in 0.18 micron technology and featured a modified elementary pixel unit that is capable of rerouting the output of any photodiode from the image sensing circuitry to a power recovery one.

2.3.2 Light emitting structures

In this subsection integrated photoemitters are discussed. The discussion will include a brief overview of photoemitting structures that have been fabricated using non-standard manufacturing, i.e. using process steps that are not normally part of any commercially available CMOS process. It is noteworthy that although significant improvements in efficiency can be achieved using these methods, the reality remains that for commercially available CMOS technology all light emitting devices that can be manufactured remain inefficient, thus imposing the restriction that a lot of power has to be dissipated on-chip before any light can be emitted[10]. This section includes a representative overview of the rich literature on the subject.

To begin with one must note once again that the fact that Silicon is an indirect bandgap material has not escaped anyone's attention. This is the starting point for a 1993 paper that seeks to review the issues related to efficient light emission from Silicon and summarise the theoretical solutions to it [103]. The solutions discussed involve the use of impurities in order to create radiative recombination centres in a material that is otherwise bereft of them, band structure engineering through the use of alloys (and for that matter the engineering of semiconductor doping), quantum confinement by use of, amongst other

[10]This can be compared to the case of optical modulators where the light source bears the burden of high power dissipation but the modulator chip operates comfortably with a very limited power budget. This has implications on how the chip can obtain the power required for its operation, the thermal profile of the chip under given environmental and operational conditions etc.

things, quantum dots, and hybrid manufacturing whereby a material that is far more amenable to light emission is grown atop Silicon. Subsequently the reader will recognise where each reference fits into this quartet of possible solutions although this literature review concentrates on the 1st couple of solution types.

2.3.2.1 Porous Silicon

Porous Silicon was one of the attempts to make simple Silicon material amenable to the emission of light. The integration of porous Silicon atop regular Silicon has been described by means of creating a comparably efficient integrated LED in standard bipolar technology (as opposed to simple crystaline Silicon) [104]. This work dates from 1996, indicating that the idea is not new. The authors quote 0.1% external power efficiency. Literature pertaining to the theoretical background of this work and earlier attempts at creating porous Silicon LEDs (not integrated though) can be found amongst the references of [104]. Amongst them this work [105] details an earlier attempt to manufacture porous Silicon in a manner much akin to integrated circuit fabrication. Meanwhile the emission spectrum of porous silicon has been shown to be centred around 650 nm [106], although this reference doesn't clarify what parameters influence either the external efficiency or the wavelength of the emitted radiation.

2.3.2.2 Erbium-doping

Doping with Erbium was another method used in order to coax Silicon into emitting light more efficiently. This practice seems to originate in work reported in 1994; work that has studied the phenomenon from a basic, physics point of view [107]. Basic properties such as wavelength of emission ($\lambda = 1.54\mu m$ under the conditions of testing such as temperature used) and emitted light intensity vs. current density were also presented in [107]. As time progressed, the Erbium-based technology entered more practical settings. Test devices were fabricated and their physics explained [108]. The lynchpin seems to revolve around careful manufacturing of the device whereby under reverse bias the Erbium atoms are part of the depletion region and therefore provide an excitation target for tunneling electrons. Thus, instead of crossing the entire depletion region and being subjected to Auger recombination[11], travelling electrons now excite Erbium atoms through impact excitation. The recombination at the Erbium sites seems to be predominantly radiative. Under forward bias, on the other hand, the Erbium hides within an ocean of free carriers and non-radiative Auger recombination dominates. The paper refers to efficiencies in the region of 2×10^{-4}, or of the order of 0.01% ('quantum efficiency' as stated in the publication itself - whether this is external or internal is not entirely clear). Finally, work to characterise Erbium-doped structures electrically has also been undertaken [109].

2.3.2.3 Dislocation-engineered Silicon

Introducing so-called 'dislocation loops' into Silicon is yet another technique for achieving electroluminescence. The idea is described in a 1996 work from the Max-Planck institute [110], where the emission of IR light from known dislocation site types (D1 and D2 type) is studied. The conclusion is that elec-

[11] Auger recombination: A process whereby an electron-hole pair recombines, but the energy is transfered to another carrier rather than resulting in the emission of a photon. The carrier then loses its newly gained extra energy to the lattice through phonon scatter. The Auger recombination rate tends to increase with doping concentration.

troluminescence from Silicon is possible at these sites, but with an external efficiency of about 1ppm. In terms of wavelength, dislocation loops seem to enhance emission at 0.78 micron and 0.88 micron wavelengths with band-to-band transitions also present in the region of 1.1 micron. The trail of dislocation loops has then been followed by [111], where the construction of a photodiode with deliberately induced dislocations within its volume led to a light emitting structure with external efficiency of the order of 100ppm (10^{-4}).

The fabrication process and physics of dislocation-engineered Silicon has also been extensively studied. For example the effects of temperature on the creation of the coveted dislocation loops [112], the optimisation of the annealing process with an eye towards creating efficient electroluminescent dies [113] and dislocation loops within the context of thermal quenching see [114] have all been studied in dedicated, highly specialised research publications. Once again external efficiencies of the order of 100ppm (in fact approx. 200ppm) are reported throughout the literature for the Erbium-doping technique. Finally, a paper reviewing the technique of engineering dislocation loops is presented here [115]. The literature continues on this subject, of course, but these publications should cover the basics.

2.3.2.4 Commercially available Silicon

At the same time people have tried working with commercially available Silicon for the purposes of photoemission and have obtained the typical efficiencies of the order of 10s of ppm. Starting from around the end of the 80's researchers have been thinking about using CMOS Silicon 'as is' for optical emission. A good starting point is this 1992 paper [116] that showcases a designed and fabricated example whilst featuring a rich literature revolving around the theory behind the application.

In the mid-90's a group comprising of, amongst others, Drs. L. W. Snyman, M. du Plessis and H. Aharoni enters the scene and creates a very wealthy literature on the subject, which will not be referenced here in its entirety. Early work by the group included the design and characterisation of a CMOS-compatible LED; an early attempt to find out roughly what sort of performance can be expected from fully integrated CMOS LEDs [117]. The quoted results are 8 nW per 20 μm diameter sector of the die for a power dissipation of 20 mW. A plot of optical power vs current, a traditional I-V curve and an optical intensity vs. wavelength plot are all provided within the same publication. Further developments in the field of CMOS-compatible LEDs eventually led to the creation of a device that can be controlled by means of applying a bias voltage at a gate terminal as opposed to merely changing the voltage across the anode and cathode of the LED and thereby modulating the reverse bias avalanche current through the device [118]. In this sense the device is more akin to a photo-transistor. Eventually the field advanced enough in order for a comprehensive study on the geometric set-up of Si LEDs both in terms of layout as seen by the designer (the geometrical patter 'as seen from above' so to speak) and in terms of fabrication (that is to say in cross-section) to be undertaken [119]. Data on electro-optical characterisation is also provided in the same publication. The emission spectrum of the studied LEDs has been shown to centre around 600nm [120]. The potential applications of Si LEDs in commercially available CMOS have also been considered [120]. After about 2004 this trail 'goes cold' as apparently the main group in this field moved on to pursuing other interests.

2.3.2.5 Other work

Finally, a couple of papers that cannot be effectively categorised in the categories above should be noted. These give an indication of other attempts to coax Silicon into emitting light. The 1st example is [121], which employs a vast array of measures intended to improve the light emission efficiency in Si, resulting in a final external efficiency of up to approx. 1%. This is achieved by mobilising semiconductor manufacturing industry to its very limits, including creating the emitter in the 'inverted pyramid' configuration (apparently to enhance light emission in a particular direction), fabricating the diode with a back-reflector and a front anti-reflective coating (again to help emission of light in a particular direction), light doping (to limit free carrier absorption of any emitted light) and float-zone wafer fabrication since this process is known to yield higher quality Silicon exhibiting higher minority carrier lifetimes i.e. lower rates of recombination. It should be noted that because of its preoccupation with 1- and 2-photon emission this paper would probably be more at home in a section about stimulated light emission from Silicon.

The 2nd example is a Silicon laser [122]. This work isn't technically concerned with electrooptics, but still it forms an interesting addition to the literature review section dealing with Silicon light emitters. The basic idea is that a laser pump feeds into a set of waveguides formed in Silicon and creates a lasing effect via ring resonation. Very 'fancy' manufacturing is involved, however it is a method that does allow the creation of Silicon lasers albeit necessitating an external pump. This paper would also rightly belong into the category of stimulated emission in Silicon.

To wrap up the section this review paper must also be mentioned [123]. It features a rather comprehensive view of light emission in Silicon, particularly at the beginning of Section 4 (of the review paper), where a brief paragraph is devoted to each section, complete with some indicative references. It must be noted that the paper also covers stimulated light emission in Silicon with an eye towards integrated lasers. These domains are outside the scope of this literature review, but can be easily pursued by looking at section 4 of the aforementioned paper and following the references therein. A table summarises this subsection of the literature review (Table 2.4).

2.3.3 Light modulation structures

This subsection deals with the modulation of light in silicon. Before any presentation of published work is done it must be clarified that modulation of light has two facets. An optical modulator may either modulate light in terms of amplitude, in which case we may seek a metric for modulation in the formula $\frac{P_{out}}{P_{in}}$, or in terms of phase, where the metric revolves around the $\Delta\phi$ between unmodulated and modulated output (phase of the output at a given detection point if the modulator is active vs. inactive). These facets dominate the world of electrooptical modulation, although it is not at all the case that they should be the only ways in which light may be subjected to modulation. In theory the polarisation, direction or wavelength of the portion of incoming light that is permitted to exit the modulating structure may also be subjected to modulation.

In order to modulate a beam in either amplitude or phase, the refractive index of the material comprising the modulator needs to change, typically as a function of some electric field applied to it. The refractive index, however, is a complex number. Changing the value of the real part of the index

Table 2.4: Summary of common methods for silicon-based light emission. The 1st column provides relevant references for each technique with efficiency quoted in the 3rd column. Footnotes indicate the exact term by which these efficiencies are described. The wavelength of emission is an indicative value since the spectrum is never at a single wavelength and even the peak will shift with temperature for distance. The last column therefore provides references containing plots of emission spectra.

Ref.	**Method**	**Efficiency**	**Emission wavelength**	**Emission spectrum**
[104]	Porous Silicon	1000 ppm*	650nm	[106]
[108]	Erbium doping	200 ppm†	1540nm	[108]
[114]	Dislocation loops	200 ppm†	780 nm, 880 nm, 1100 nm, 1600 nm, etc.	[110, 112] 114,
[117]	Plain SiLED	2 ppm in [117], but typ. $\leq$0.1 ppm*	650nm avalanche, other wavelengths too far field emission	[117]
[121]	Inverted pyramid topology	10000 ppm‡	1150nm	[121]

*External power efficiency, †Quantum efficiency, ‡Power conversion efficiency.

causes the phenomenon of electrorefraction to play its role by modulating phase, whilst changing the value of the imaginary part leads to the phenomenon of electroabsorption coming into force and modulating amplitude.

Nevertheless, in this work the focus lies solely on amplitude/phase modulation, as this very type of modulation is what has been used to achieve the aims of the project. Note: electro-optical effects in Silicon are many and varied, however the bulk of the literature below is dominated by the free carrier absorption and refraction phenomena. Since the presence of free carriers causes both amplitude and phase modulation in free carrier-imbued semiconductors, amplitude and phase modulation are inextricably intertwined with one another, hence the term 'phase/amplitude modulation'. The project described in this thesis only uses the amplitude modulation branch of the amplitude/phase modulation. The bulk of the literature tends to rely mostly on the phase modulation branch of the duo.

To this end, the free-carrier absorption and refraction effects are used to achieve modulation in Silicon. An early reference on the subject can be found here [124]. The authors set-up the basic theory for the phenomenon of free carrier (FC) absorption and follow-up with measurements. Next, a very widely cited, key reference by Drs. Soref and Benett laid out a lot of groundwork by studying both phenomena (absorption and refraction although the stress is on the latter) both theoretically and with measurements and quantifying them for use in various applications [125]. Plots within this paper show the theoretically expected and measured changes in refractive index (separate plots for the real and imaginary parts) as a function of free carrier concentration whilst equations describe the phenomena from a theoretical point of view. Even though the theory shows a discrepancy of a factor < 10 in comparison to the obtained results, this discrepancy seems to be consistent, indicating that the theory describes the phenomenon fairly well, at least from a qualitative point of view. An improvement over the basic Soref/Benett theory can be found in this paper from Fraunhofer Institute [126].

The detailed physics of how free carrier absorption arises and how it was used throughout this project

can be found in the background and theoretical chapters of this thesis (chapters 3 and 4 respectively).

2.3.3.1 Carrier injection approach

Based on this theory a few approaches have been used in order to exploit the free carrier absorption and refraction phenomena (we'll just call them FCPs for Free Carrier Phenomena hence forth). To begin with, the simplest approach is to create a semiconductor structure which is normally mostly devoid of free carriers, but can be flooded with large numbers of them under the right circumstances (i.e. upon command from a control signal). To achieve this, one may rely on carrier injection (see Figure 2.7 (a)). To this end there a fairly rich literature of which only a few specimens will be presented. First, work involving Soref in 1987 [127] showing an early example of a device achieving optical switching based on carrier injection.

The problem with carrier injection was then found to be that the carrier lifetime would be a strong limiting factor in what pertains to the speed of any such device. However, this was overcome when carrier injection-based technology met micro-resonators (for extra information on resonators please see the 'additional' sub-subsection following). To illustrate the concept of a microresonator one of the 1st micrometer scale resonators in Silicon can be found in [128]. Here, carrier injection is introduced as a method to achieve control of the free carrier concentration in the optical channel and the problem of carrier lifetime is mentioned. Nevertheless, with higher voltage driving and helped by the resonating nature of the modulator the authors achieve 1.5 Gbps data transfer rates, which are close to and in fact slightly above the theoretical limit for carrier injection devices (non-resonant as described in [129]). Next, the combination of these technologies is refined by the introduction of the 'pre-emphasis' driving scheme whereby a short, but high V pulse is sent to the modulator control terminal before the actual signal arrives. This 'pre-emphasis' allows the system to be shocked into a much more responsive state just before the signal actually arrives. This breakthrough allowed the inventing team to achieve modulation of first 12.5 Gpbs [130], and then 18 Gbps [131]. The former of these papers is where a description of the pre-emphasis signalling technique can be found.

2.3.3.2 BMFET technique

A theoretical and simulated study of a device that inject carriers into the region of the optical channel, but then instead of attempting to shuttle them in and out of the system entirely, merely shuttles them in and out of the optical path while keeping them in the general vicinity has been presented [132] (see Figure 2.7 (b)). Proper biasing of certain control terminal(s) can achieve this. The paper explains the method thoroughly and provides, amongst other things, a table where a comparison of optical modulators is performed covering work between 1987 and 1996. Follow-up work describes the mechanism again, but concentrates more on the underlying mechanics and fabrication of the device, which is called a 'BMFET' (Bipolar Mode Field Effect Transistor) [133].

2.3.3.3 Depletion of carriers method

Devices based on depletion mode (see Figure 2.7 (c)) were a method designed to break free from the limitations of carrier lifetimes that plague carrier injection devices, as reported originally in [134]. The

results are merely simulated, but the idea is set. By 2007 the system had been optimised to transmit up to 40 Gbps [135] and by 2012 to 50 Gbps [136].

The literature in this subject is very rich since researchers have created countless semiconductive structures meant to optimise modulation in Silicon. These typically involve a number of interdigitated p- and n-type slices of semiconductive material of varied doping concentrations. Examples abound. On one occasion a heavily doped slit has been added within the optical path in order to add extra contrast between normal and depleted states (greater difference in carrier concentrations) [137]. A more advanced example where smaller waveguides, increased doping cocnentration and optimised localsiation of the metallurgical surface of the pn-junction are mobilised in order to improve modulation efficiency [138]. An example depletion mode system for use in ring resonators can be found in [139]. Alternatively a brief review of these structures can be found in [140].

2.3.3.4 Charge accumulation approach

A 4th method for building an FCP-based modulator relies on charge accumulation. This approach is based upon the presence of a dielectric material within the optical path. This allows charged carriers to be deposited either side of the dielectric material when a bias voltage is applied across it, exactly as it would in any dedicated capacitor structure. The carriers can then be emptied by use of appropriate biasing of this 'optoelectronic capacitor' for use of a better term (see Figure 2.7 (d)). Once again, carrier lifetime doesn't interfere with this mechanism. Intel pioneered this approach and reported this in [141]. Other work following this direction, as for example [142], where the geometry of the accumulation zone is made more elaborate, or [143], where an 8 mm-by-1,5 μm optical modulator is presented.

2.3.3.5 Related work

As a sideline a loosely related idea where a combined power harvester and modulator device is designed can be found in [144]. The modulator is based on carrier injection and external biasing to switch between the ON and OFF states (that is to say switch the modulation ON or OFF) while staying entirely in 4th quadrant operation, thereby being a net energy supply device throughout. Perhaps this device could find use in applications where power dissipation is far more critical than speedy signal recovery.

Finally a set of review papers will form a good starting point for anyone who wishes to study the field of Silicon modulators in more depth. Please note that this literature review has not covered hybrid systems such as those where Ge, organic compounds or other substances are used in tandem with Silicon in order to improve the quality or speed of modulation. The review papers can be summarised as such:

- A publication looking towards the future of optoelectronics in Silicon, written by Soref himself in 1987. It is more of historical value, although the physical principles summarised therein haven't changed [145].
- An explanation of the basic principles of operation behind electrooptic modulation in Silicon as they have been applied by various groups, details the important metrics used in characterising the modulators and summarises the state of the art as it appeared in 2010 when it was written [129].
- A review of various types of modulators with very helpful cross-sections of each of them [146].

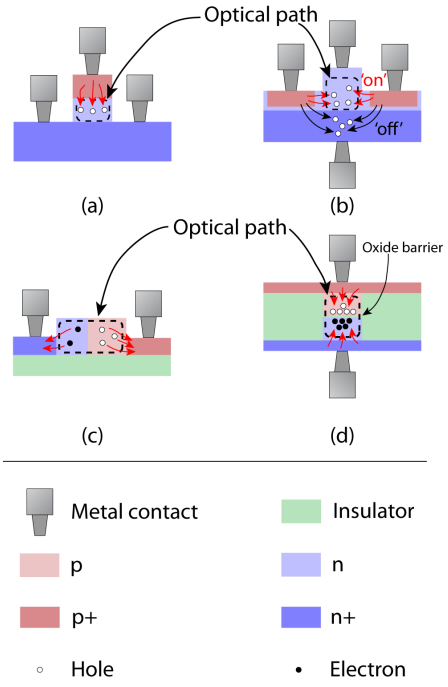

Figure 2.7: Mainstream modulators in waveguide-based Silicon photonics. Cross-sections of the devices (sections perpendicular to the direction of light travel) are shown with dashed outlines showing the cross-sections of the beams to be modulated in each case. a) Carrier injection mode. The arrows show free holes being injected into lightly doped n-type material in order to increase free carrier refraction. b) BMFET (shifting carriers in and out of the optical path). Arrows show 'on' and 'off' states. c) Carrier depletion mode. Arrows show carriers leaving lightly doped semiconductor regions in order to reduce free-carrier refraction. d) Charge accumulation mode. Arrows show free carriers being accumulated around the insulating barrier, which lies within the optical path.

Additional

A large number, in fact the vast majority, of the modulators described above were intended for use within the context of larger optoelectronic blocks. The most popular such blocks are ring (or racetrack) resonators and Mach-Zender interferometers. Very briefly:

1. Resonators are based on manufacturing a semiconductor waveguide that is intended to transmit the signal in very close proximity to a looped semiconductor that is intended to host a standing wave pattern. By changing the refractive index of the ring through electro-optical effects the resonance frequency shifts. The effective optical coupling between the main waveguide and the looped resonator is significantly different at the resonant frequency than at other frequencies, which is why at the output of the waveguide the resonance wavelength of the cavity can be detected either as a 'notch' in the output spectrum [12] or as a 'bump' depending on the specifics of the resonator configuration (the setup may involve more loops and more signal waveguides in more advanced cases). A review on resonating systems can be found here [147].
2. Mach-Zender interferometers are devices intended to measure the difference in phase between a couple of beams. Of course the light needs to be coherent. Typically what happens is that a beam of coherent light is somehow split into 2 branches (achievable by half-silvered mirrors for example or by simply capturing different sub-fascicles of the whole beam fascile). Then the branches are routed through different pathways which include modulators. The modulators presented above would fit into these parts of the optical path. After being subjected to different degrees of phase modulation (and willingly or not also amplitude modulation for FCP-based modulators) the beams are recombined and sent onwards for futher processing. At that stage it is possible to measure the phase difference between the now merged beams.

2.4 Power management in optical systems

Optical power entering a system from a photo-electric element is almost invariably characterised by the very low voltage at which it manifests itself, which is why DC-DC conversion systems are necessary. They operate by ramping up the voltage to a level that can be subsequently down-regulated to a stable power supply with enough headroom to accommodate a variety of circuits. The choice of circuit topology, but also the types of devices available in the manufacturing technology will play a role in determining the efficiency and ultimately fitness-for-purpose of the implemented DC-DC converter.

Furthermore because a pn-junction-based photo-electric element is basically a diode in parallel to a current source (+ a few resistors) as seen in Figure 2.8, there is a real danger that a PMU may allow too much energy to flow through the diode (one of the parasitic loss paths - see Figure 2.8) instead of capturing it and supplying it to the load. Specifically if the voltage across the pn-junction is allowed to increase significantly, then on the positive side the PMU's devices will work faster and the unit as a whole will work more efficiently as the transistors therein are more likely to be driven properly at their gates, but on the negative side the diode will start passing unsettlingly large amounts of current. The reverse occurs when the voltage across the diode is kept low. For that reason it is mathematically imperative

[12] By this it is meant that the the output spectrum is almost identical to the input spectrum except for a 'notch' appearing at the ring resonating frequency in the output spectrum.

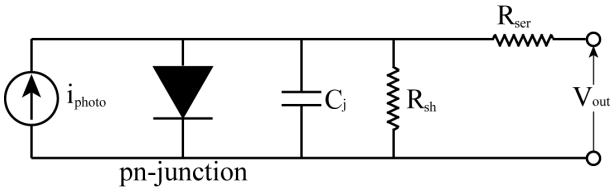

Figure 2.8: Macromodel of a pn-junction photoelectric element. The current source represents electron-hole pair generation by incoming light. The diode structure represents the main electrical part of the element with non-idealities modelled through the junction capacitor (C_j), the shunting resistor (R_{sh}) and the series resistors (R_{ser}). Parasitic loss paths involve the shunting resistor and for higher frequency components of the photocurrent also the junction capacitance, however both of these are negligible compared to the losses through the diode under typical operating conditions.

that an optimum bias point exists in between these limits whereby the PMU is providing the maximum amount of energy to the load. Circuits designed to ensure that a given PMU can always operate close to that point employ MPPT (Maximum Power Point Tracking) systems.

This section of the literature review will deal with some of the main constituent parts of PMUs: (a) Charge pumps. (b) MPPT systems.

Of course power management units will typically include other elements, such as a ring oscillator to generate some sort of clock signal or start-up circuits to ensure function at the correct operating point, voltage references and output regulators. The literature on each subject listed above and circuit elements such as oscillators and start-ups is extremely extensive. A resource covering all mentioned subjects can be found in this book on CMOS circuit design [148].

2.4.1 Charge pumps

When it comes to charge pumps, there is a 2010 review paper that goes through their principles of operation and lists the most popular topologies [149]. According to the this source, the main topologies are: Dickson, MOS version of Dickson, Bootstrap charge pumps and their doubles, latched, series-parallel and adaptive charge pumps. To those, one could add the Pelliconi cascade and the Makowski cell.

2.4.1.1 Dickson-derived designs

The commonly thought of as most basic version of a charge pump originates from Dickson [150]. It was a simple design from the days when components were discrete and power supplies high and revolved around capacitors and diode-based rectifiers (see Figure 2.9). The paper is still often cited but as a topology it has been more or less abandoned in favour of more efficient developments. Nevertheless, many of these developments are very similar in spirit to the Dickson topology which is why it remains a relevant reference today. Finally, research published in these references [151, 152, 153] carries on Dickson's work by modeling the operation of the charge pump at increasing levels of complexity.

The MOS version of the Dickson charge pump is merely the same configuration whereby the diode elements have been replaced by diode-connected MOS transistors.

Logically the next step up from the Dickson pump would be the Makowski cell (although intermediate

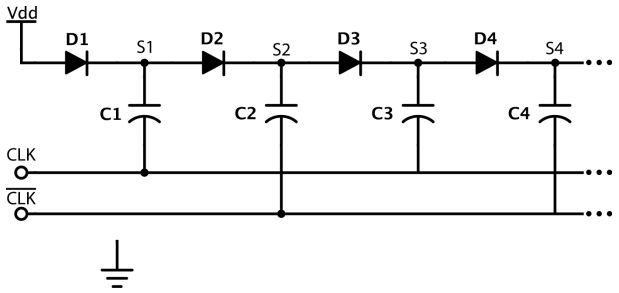

Figure 2.9: Dickson charge pump concept. Power arrives from node 'Vdd' and is subsequently boosted at each stage (S1, S2, ...) under the control of the non-overlapping clocks CLK and $\overline{CLK}$. See [150] for details.

stages may have occurred). In a highly mathematical paper Makowski explained the theoretical basis behind the topology [154]. A design strategy based on analysis of the Makowski charge pump was then developed and used in the design of a 4 stage and a 6-8 stage Makowski cell [155].

The next step up would be the bootstrapped charge pumps. According to [149] the origins of the idea are to be found in [156, 157] although the schematic associated with the concept doesn't appear in either of those publications as it appears in the review paper. This topology relies on clock boosting in order to minimise losses for charge transfer between the stages of the charge pump. The objective is achieved, but in exchange a quad-clock regime is required with proper phase control for each component.

The double versions of these circuits are simply pairs of each topology working in parallel. The problem with single charge pumps is that they work on a clocked regime and only provide power for 1/2 of the duty cycle while using the remaining 1/2 to charge up. For that reason pairs of charge pumps working in anti-phase regime can provide a much more stable power supply.

The next set of charge pumps derive mostly from the Pelliconi cell[13], as seen in [158, 159]. Though there is earlier work featuring this topology Pelliconi's group generated the aforementioned dedicated paper to study it. This line of research was followed by others such as in this paper [160], where the full Pelliconi cascade is studied as a chain.

2.4.1.2 Series-parallel approach

Yet another charge pump technique involves the so-called 'series-parallel' approach (a.k.a. switch-capacitor based charge pumps), whereby capacitors are connected first in parallel to the low voltage supply (charging phase) and then connected in series to boost up the voltage (discharge phase). An example of such system can be found here [161]. A nice comparison between certain series-parallel topologies vs. the standard Dickson or Makowski configurations can be found here [162].

Finally, adaptive charge pumps are a different flavour of series-parallel pumps where the switching pattern can be adapted in order to configure the same circuit into behaving as a charge pump of different numbers of stages. A practical example can be found here [163].

[13] Pelliconi and Dickson topologies are similar in spirit, which is they are all grouped under the Dickson-derived designs.

2.4.1.3 Other approaches

There have been other approaches in boosting voltage, a popular one being developed in [164]. The concept relies on using actively biased transistors, rather than essentially diode-connected devices. Capacitors connected to clocked inputs (much like in the Dickson design -see Figure 2.9-) are allowed to charge during the low phase of the clock and then discharge at the high phase under the control of actively biased switches. In [164] the active biasing is provided by cross-coupling two such switches and operating them in an interleaved fashion (see publication for more information).

This basic design concept has given rise to derivatives such as the ones in [165, 166] and has been the subject of analysis, as in [167]. We only briefly mention the existence of such systems since it is difficult to generalise beyond voltage doubling (i.e. to $\times N$ voltage multiplication) and will typically be used for applications where the initial voltage is higher than what is traditionally provided by optical power harvesters.

2.4.2 Maximum power point tracking (MPPT) systems

Again, implementing an MPPT and innovating in the area was outside the scope of this thesis, however this subsection has been included for completeness.

A number of review papers can be used as guides to the subject of MPPT systems. An example that lists the types of MPPT systems ranging from classical 'hill-climbers' and 'perturb & observe' to neural and fuzzy systems can be found in [168]. A point is made that MPPT systems come in a large variety of implementations. MPPT systems of a more 'circuity' flavour are examined in [169]. Particular attention is paid to curve-fitting, look-up table, open circuit voltage-based, short circuit current-based, direct sampling and $\frac{dP}{dt}$ methods. These methods are more oriented towards circuit implementations in contrast to more software-heavy artificial intelligence-based or fuzzy systems. More recent work (2011) reviews more advanced types of MPPT systems, some revolving around complicated control algorithms [170] whilst a comparative study of 4 types of MPPT systems was presented in [171].

Finally, an example of a very well-established hill-climbing MPPT system can be seen unfolded in [172]. The reference includes schematics for all system components.

Note: Many of the systems mentioned in the literature above are intended for implementation on large solar arrays where the power cost of the controller system is not as critical as it would be on power-scavenging applications where both solar cell, PMU and MPPT reside on the same die and share the same -very limited- resources. As such, some of the more computation- and hardware-intensive MPPT varieties are less likely to be found in integrated circuits, but simpler approaches such as hill-climbing and perturb & observe do often appear with schematics for IC integration even though certain components may sometimes be left off-chip (e.g. [173]).

2.5 Optical transceivers

This final section of this chapter discusses optical telemetric systems designed as a whole. Here we tackle systems that communicate chip-to-board style, as opposed to intra-chip waveguide-based systems. The

links to the previous couple of sections are evident. Some general remarks on optical transceivers are made upon the opportunity.

To begin with, [174] gives an overview of optical interconnects as seen in 2001. The case is made that electrical interconnects can become bottlenecks and that the issue of designing fast and power-efficient interconnect systems is not to be trifled with. The issue of monolithic integration of all optoelectronic elements is also discussed as well as hybrid solutions where Silicon dies and compound semiconductor dies are bonded together with the Silicon handling the processing and the compounds dealing with the optoelectronics. Next, a brief overview of optoelectronic devices is presented and optics-friendly packaging is discussed. It must be noted that for most systems this implies packaging that can host waveguides through which light can propagate much akin to current through a wire (although, alas, with much less success in bending around corners).

The nature of a typical inter-chip optoelectronic communications system can be seen in both conceptual diagrams and in actual Silicon in [175]. Figure 1 of the aforementioned reference illustrates the concept and makes it clear that the optical parts of the system are not meant to be integrated into Silicon. VCSELs (Vertical Cavity Surface-Emitting Lasers) and dedicated photodiodes are tasked with generating and detecting light (entirely off-chip) whilst optic fibres carry the onus of transmitting the light generated by VCSELs to their respective detectors. The Silicon part of the set-up is there to drive the VCSELs and decode the incoming photocurrent from the photodetectors. The reference also includes descriptions of the circuits used for these functions. Later on a diagram (Figure 20) of the optics involved in the system is provided (VCSEL to lens assembly, to optic fibre, to lens assembly to photodetector). A demonstrator system is then presented and characterised.

The part of the optical path that lies outside chips is also important when it comes to inter-chip communication. Work discussing precisely this aspect and proposes various optomechanical components that would help integrate dies with optoelectronic capabilities into a larger framework, much like conventional PCBs integrate fully electronic ICs into a larger framework can be seen in [176].

A full CMOS-compatible optoelectronic transceiver with an integrated light source has been created [177]. The Silicon emitter is weak, as would be expected of it, but despite the modest speed of the system (176 kbps quoted linkage speed), it can be called genuinely CMOS compatible as none of the manufacturing seems to have required any special processing steps, not even the primitive reflector structure meant to help couple stray light into the fibre (at least as far as can be deduced from the paper - intricate details of the manufacturing aren't provided). Given the difficulties encountered in achieving decent light emission from Silicon even this result is no small achievement. The break from the common wisdom that optoelectronic platforms need to either employ off-chip specialised optoelectronic structures or utilise hybrid semiconductors (III-V patches on Silicon for example) is noteworthy in itself.

Finally some comparative data between inductive, optical and wired systems can be seen in [178]. The graphs in figure 18 of the reference show delay, energy/bit and bandwidth density vs process node for standard bus, inductive and optical interconnects. The metrics are compared between a 2cm link that is either implemented in wire, or via inductive coupling, or optically. Further specifications about how the values in the aforementioned graphs have been calculated and exactly to what sort of systems they pertain to can be found within the text of the reference (for example the wired interconnect is based on a parallel repeater bus architecture and the RF link on the RF-I technology). In terms of delay inductive (RF) and optical coupling show approximately similar delays (around 250 ps) for all technology nodes

(90nm through to 22nm) vs. values in the region of 1500 ps for wired interconnects. In terms of energy per bit RF and optical links achieve around 1 pJ/bit vs. 11-17 obtained by the bus. Notably, as technology scales down the figures for both the RF and the bus links improve whilst the corresponding number for the optical link remains constant. Finally the bandwidth density picture changes dramatically between the 90nm and the 22nm node with: a) The RF link performing better than either bus or optical links at 90nm. b) RF and optical links reaching parity at the 22nm node while the bus link lags behind. c) Constant improvement for all methods with lower feature size technology. These figures hint towards the conclusion that optical and RF linkage systems are set to remain targets of active research as they promise to deliver fast and energy-efficient communication over what by microelectronic standards is considered 'long distance'.

2.6 Summary

By now the intricacies and innumerable facets of optoelectronic communication system design have been reviewed by taking a look at the related literature. This thesis adds a new perspective in the domain of leveraging optoelectronics for the purposes of communications. Departing from the typical systems seen today that use dedicated off-chip optoelectronic devices we attempt to create all the elements that a wire-bond-free integrated circuit will need in order to be able to handle all data transactions without the help of any off-chip components. Our concept uses on-chip light modulation instead of on-chip light generation for the purpose of sending a signal to the outside world. Data read-in and power delivery will be tackled using traditional integrated photodiodes whilst for the generation of a stable power supply a simple, but rugged power management unit will be used.

Bibliography

[1] A. C. Bowman, "A selective encapsulation solution for packaging an optical micro electro mechanical system," Master's thesis, Worcester Polytechnic Institute.

[2] U. S. patent office, "Us patent 6.489.178 b2." patent, December 2002.

[3] C. Cotofana, A. Bossche, P. Kaldenberg, and J. Mollinger, "Low-cost plastic sensor packaging using the open-window package concept," *Sensors and Actuators A: Physical*, vol. 67, no. 13, pp. 185 – 190, 1998.

[4] A. Bos, L. Wang, and T. van Weelden, "Encapsulation of the next generation advanced mems and sensor microsystems," pp. 1 –5, june 2009.

[5] W. Oelssner, J. Zosel, U. Guth, T. Pechstein, W. Babel, J. Connery, C. Demuth, M. G. Gansey, and J. Verburg, "Encapsulation of ISFET sensor chips," *Sensors and Actuators B: Chemical*, vol. 105, no. 1, pp. 104 – 117, 2005. Piet Bergveld Special Issue.

[6] T. Velten, M. Biehl, W. Haberer, T. Koch, P. Ortiz, N. Keegan, J. Spoors, J. Hedley, and C. McNeil, "Packaging of a silicon-based biochip," pp. 720603–720603–10, 2009.

[7] C. Harper, *Electronic Packaging and Interconnection Handbook*. New York, NY, USA: McGraw-Hill, Inc., 4 ed., 2005.

[8] H. Ruf, T. Knoll, K. Misiakos, R. Haupt, M. Denninger, L. Larsen, P. Petrou, S. Kakabakos, E. Ehrentreich-Frster, and F. Bier, "Biochip-compatible packaging and micro-fluidics for a silicon opto-electronic biosensor," *Microelectronic Engineering*, vol. 83, no. 49, pp. 1677 – 1680, 2006.

[9] J. E. Shaw, "Capillary fill encapsulation of ISFETs," *Sensors and Actuators A: Physical*, vol. 3738, no. 0, pp. 74 – 76, 1993. Proceedings of Eurosensors VI.

[10] T. Prodromakis, P. Georgiou, T. Constandinou, K. Michelakis, and C. Toumazou, "Batch encapsulation technique for CMOS based chemical sensors," in *Biomedical Circuits and Systems Conference, 2008. BioCAS 2008. IEEE*, pp. 321 –324, nov. 2008.

[11] L. Li and A. Mason, "Post-CMOS parylene packaging for on-chip biosensor arrays," in *Sensors, 2010 IEEE*, pp. 1613 –1616, nov. 2010.

[12] J. M. e. a. Rothberg, "An integrated semiconductor device enabling non-optical genome sequencing," *Nature*, March 2010.

[13] E. Ghafar-Zadeh, M. Sawan, and D. Therriault, "A microfluidic packaging technique for lab-on-chip applications," *Advanced Packaging, IEEE Transactions on*, vol. 32, pp. 410 –416, may 2009.

[14] R. Dudek, D. Vogel, and B. Michel, "Mechanical failure in cob-technology using glob-top encapsulation," *Components, Packaging, and Manufacturing Technology, Part C, IEEE Transactions on*, vol. 19, pp. 232 –240, oct 1996.

[15] S. Hian, F. C. Seng, T. S. Kiong, and N. Kalandar, "Wire diameter and length effects on wire sweep performance of multi-tier copper and gold wire bonding in plastic ball grid array packages," in *Electronics Packaging Technology Conference (EPTC), 2011 IEEE 13th*, pp. 764 –768, dec. 2011.

[16] L. Lantz, S. Hwang, and M. Pecht, "Characterization of plastic encapsulant materials as a baseline for quality assessment and reliability testing," *Microelectronics Reliability*, vol. 42, no. 8, pp. 1163 – 1170, 2002.

[17] W. Fan, Y. Jeong, J. Wei, B. Tan, B. Lok, and K. Chun, "Encapsulation and packaging of biosensor," in *Electronic Packaging Technology Conference, 2005. EPTC 2005. Proceedings of 7th*, vol. 1, p. 4 pp., dec. 2005.

[18] P. Hammond and D. Cumming, "Encapsulation of a liquid-sensing microchip using su-8 photoresist," *Microelectronic Engineering*, vol. 7374, no. 0, pp. 893 – 897, 2004. Micro and Nano Engineering 2003.

[19] O. Brand and H. Baltes, "Microsensor packaging," *Microsystem Technologies*, vol. 7, pp. 205–208, 2002. 10.1007/s005420100110.

[20] H. Baltes and O. Brand, "CMOS-based microsensors and packaging," *Sensors and Actuators A: Physical*, vol. 92, no. 13, pp. 1 – 9, 2001. Selected Papers for Eurosensors XIV.

[21] H. Baltes, O. Brand, and M. Waelti, "Packaging of CMOS mems," *Microelectronics Reliability*, vol. 40, no. 810, pp. 1255 – 1262, 2000. Reliability of Electron Devices, Failure Physics and Analysis.

[22] C. H. et al., "Packaging design for implantable microstimulator," *Journal of medical and biological engineering*, vol. 23, no. 3, 2003.

[23] S. Adamson and C. Ness, "Dam & fill encapsulation for microelectronic packages," in *NEPCON WEST*, pp. 1373–1388, CAHNERS EXPOSITION GROUP, 1999.

[24] J. N. Y. A. et al., "Brain silicon interface for high-resolution in vitro neural recording," *IEEE transactions on biomedical circuits and systems*, vol. 1, March 2007.

[25] S. B. Prakash and P. Abshire, "On-chip capacitance sensing for cell monitoring applications," *Sensors Journal, IEEE*, vol. 7, pp. 440 –447, march 2007.

[26] D. Dudenbostel, K.-L. Krieger, C. Candler, and R. Laur, "A new passive CMOS telemetry chip to receive power and transmit data for a wide range of sensor applications," in *Solid State Sensors and Actuators, 1997. TRANSDUCERS '97 Chicago., 1997 International Conference on*, vol. 2, pp. 995 –998 vol.2, jun 1997.

[27] R. Catansecu and K. L. K. et al., "On-chip coils produced by electroplating," *ACTUATUR 96*, June 1996.

[28] K. L. K. et al., "Design of monolithic integrated mico-coils for wireless energy transmission," *2nd international symposium micro total analyse systems*, November 1996.

[29] J. Von Arx and K. Najafi, "A wireless single-chip telemetry-powered neural stimulation system," in *Solid-State Circuits Conference, 1999. Digest of Technical Papers. ISSCC. 1999 IEEE International*, pp. 214 –215, 1999.

[30] M. Ghovanloo and K. Najafi, "A high-rate frequency shift keying demodulator chip for wireless biomedical implants," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, vol. 5, pp. V–45 – V–48 vol.5, may 2003.

[31] M. Ghovanloo, K. Beach, K. Wise, and K. Najafi, "A biCMOS wireless interface chip for micromachined stimulating microprobes," in *Microtechnologies in Medicine amp; Biology 2nd Annual International IEEE-EMB Special Topic Conference on*, pp. 277 –282, 2002.

[32] T. Kitroda, "Non-contact inter-chip data communications technology for system in a package," in *Solid-State and Integrated Circuits Technology, 2004. Proceedings. 7th International Conference on*, vol. 2, pp. 1347 – 1352 vol.2, oct. 2004.

[33] D. Mizoguchi, Y. Yusof, N. Miura, T. Sakura, and T. Kuroda, "A 1.2gb/s/pin wireless superconnect based on inductive inter-chip signaling (iis)," in *Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International*, pp. 142 – 517 Vol.1, feb. 2004.

[34] P. Irazoqui-Pastor, I. Mody, and J. Judy, "In-vivo eeg recording using a wireless implantable neural transceiver," in *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*, pp. 622 – 625, march 2003.

[35] N. Miura, D. Mizoguchi, T. Sakurai, and T. Kuroda, "Cross talk countermeasures in inductive inter-chip wireless superconnect," in *Custom Integrated Circuits Conference, 2004. Proceedings of the IEEE 2004*, pp. 99 – 102, oct. 2004.

[36] M. Ahmadian, B. Flynn, A. Murray, and D. Cumming, "Miniature transmitter for implantable micro systems," in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 4, pp. 3028 – 3031 Vol.4, sept. 2003.

[37] A. Shameli, A. Safarian, A. Rofougaran, M. Rofougaran, and F. De Flaviis, "An rfid system with fully integrated transponder," in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2007 IEEE*, pp. 285 –288, june 2007.

[38] J. Tompson, A. Dolin, and P. Kinget, "2.6-ghz rf inductive power delivery for contactless on-wafer characterization," in *Microelectronic Test Structures, 2008. ICMTS 2008. IEEE International Conference on*, pp. 175 –179, march 2008.

[39] S. Luan, A. Eftekhar, O. Murphy, and T. Constandinou, "Towards an inductively coupled power/data link for bondpad-less silicon chips," in *Circuits and Systems (ISCAS), 2011 IEEE International Symposium on*, pp. 2597 –2600, may 2011.

[40] M. Usami, "An ultra small rfid chip: mu;-chip," in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2004. Digest of Papers. 2004 IEEE*, pp. 241–244, 2004.

[41] J. Von Arx and K. Najafi, "On-chip coils with integrated cores for remote inductive powering of integrated microsystems," in *Solid State Sensors and Actuators, 1997. TRANSDUCERS '97 Chicago., 1997 International Conference on*, vol. 2, pp. 999 –1002 vol.2, jun 1997.

[42] K. Kanda, D. Antono, K. Ishida, H. Kawaguchi, T. Kuroda, and T. Sakurai, "1.27gb/s/pin 3mw/pin wireless superconnect (wsc) interface scheme," in *Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC. 2003 IEEE International*, pp. 186 – 487 vol.1, 2003.

[43] R. Drost, R. Hopkins, R. Ho, and I. Sutherland, "Proximity communication," *Solid-State Circuits, IEEE Journal of*, vol. 39, pp. 1529 – 1535, sept. 2004.

[44] J. Wu, V. Quinn, and G. H. Bernstein, "Simple wireless powering scheme for mems devices," pp. 43–52, 2001.

[45] J. Wu, V. Quinn, and G. H. Bernstein, "Powering efficiency of inductive links with inlaid electroplated microcoils," *Journal of Micromechanics and Microengineering*, vol. 14, no. 4, p. 576, 2004.

[46] E. Song, J. Kim, and J. Kim, "Multi-helix inductor of wireless power transfer system for 3-d stacked package," in *Microwave Workshop Series on Innovative Wireless Power Transmission: Technologies, Systems, and Applications (IMWS), 2012 IEEE MTT-S International*, pp. 43 –46, may 2012.

[47] Y. Y. et al., "Simultaneous 6-gb/s data ad 10-mw power transmission using nested clover coils for non-contact memory card," in *IEEE journal of solid-state circuits*, vol. 47, October 2012.

[48] T. Kuroda, "Power reduction in high-speed inter-chip data communications," in *ASIC, 2005. ASICON 2005. 6th International Conference On*, vol. 1, pp. 1145 –1149, oct. 2005.

[49] D. Mizoguchi, N. Miura, M. Inoue, and T. Kuroda, "Design of transceiver circuits for nrz signaling in inductive inter-chip wireless superconnect," in *Integrated Circuit Design and Technology, 2005. ICICDT 2005. 2005 International Conference on*, pp. 59 – 62, may 2005.

[50] N. Miura, D. Mizoguchi, M. Inoue, T. Sakurai, and T. Kuroda, "A 195-gb/s 1.2-w inductive inter-chip wireless superconnect with transmit power control scheme for 3-d-stacked system in a package," *Solid-State Circuits, IEEE Journal of*, vol. 41, pp. 23 – 34, jan. 2006.

[51] K. Niitsu, S. Kawai, N. Miura, H. Ishikuro, and T. Kuroda, "A 65 fj/b inductive-coupling inter-chip transceiver using charge recycling technique for power-aware 3d system integration," in *Solid-State Circuits Conference, 2008. A-SSCC '08. IEEE Asian*, pp. 97 –100, nov. 2008.

[52] N. Miura, T. Shidei, Y. Yuxiang, S. Kawai, K. Takatsu, Y. Kiyota, Y. Asano, and T. Kuroda, "A 0.7v 20fj/bit inductive-coupling data link with dual-coil transmission scheme," in VLSI *Circuits (*VLSI*C), 2010 IEEE Symposium on*, pp. 201 –202, june 2010.

[53] M. Ghovanloo and K. Najafi, "Fully integrated wideband high-current rectifiers for inductively powered devices," *Solid-State Circuits, IEEE Journal of*, vol. 39, pp. 1976 – 1984, nov. 2004.

[54] R. Saw and S. Jamuar, "Design of a 1.8v on-chip voltage generator for applications in low voltage transceiver," in *Circuits and Systems, 2008. APCCAS 2008. IEEE Asia Pacific Conference on*, pp. 676 –679, 30 2008-dec. 3 2008.

[55] J. Essel, D. Brenk, J. Heidrich, H. Reinisch, G. Hofer, G. Holweg, and R. Weigel, "Highly efficient multistandard rfids enabling passive wireless sensing," in *Microwave Conference, 2009. APMC 2009. Asia Pacific*, pp. 2228 –2231, dec. 2009.

[56] F. Mounaim and M. Sawan, "Integrated high-voltage inductive power and data-recovery front end dedicated to implantable devices," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 5, pp. 283 –291, june 2011.

[57] N. Miura, D. Mizoguchi, Y. Yusof, T. Sakurai, and T. Kuroda, "Analysis and design of transceiver circuit and inductor layout for inductive inter-chip wireless superconnect," in VLSI *Circuits, 2004. Digest of Technical Papers. 2004 Symposium on*, pp. 246 – 249, june 2004.

[58] N. Miura, D. Mizoguchi, T. Sakurai, and T. Kuroda, "Analysis and design of inductive coupling and transceiver circuit for inductive inter-chip wireless superconnect," *Solid-State Circuits, IEEE Journal of*, vol. 40, pp. 829 – 837, april 2005.

[59] S. Sonkusale and Z. Luo, "A wireless data and power telemetry system using novel bpsk demodulator for non-destructive evaluation of structures," in *Sensors, 2007 IEEE*, pp. 300 –303, oct. 2007.

[60] N. Miura, Y. Kohama, Y. Sugimori, H. Ishikuro, T. Sakurai, and T. Kuroda, "An 11gb/s inductive-coupling link with burst transmission," in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, pp. 298 –614, feb. 2008.

[61] J. Xu, J. Wilson, S. Mick, L. Luo, and P. Franzon, "2.8 gb/s inductively coupled interconnect for 3d ics," in VLSI *Circuits, 2005. Digest of Technical Papers. 2005 Symposium on*, pp. 352 – 355, june 2005.

[62] T. Kuroda and N. Miura, "Perspective of low-power and high-speed wireless inter-chip communications for sip integration," in *Solid-state device research conference ESSDERC*, pp. 19–21, September 2006.

[63] Y. Yuxiang, N. Miura, S. Imai, H. Ochi, and T. Kuroda, "Digital rosetta stone: A sealed permanent memory with inductive-coupling power and data link," in VLSI *Circuits, 2009 Symposium on*, pp. 26 –27, june 2009.

[64] M. Saito, N. Miura, and T. Kuroda, "A 2gb/s 1.8pj/b/chip inductive-coupling through-chip bus for 128-die nand-flash memory stacking," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, pp. 440 –441, feb. 2010.

[65] N. Miura, K. Kasuga, M. Saito, and T. Kuroda, "An 8tb/s 1pj/b 0.8mm2/tb/s qdr inductive-coupling interface between 65nm CMOS gpu and 0.1 and dram," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, pp. 436 –437, feb. 2010.

[66] T. Takeya, L. Nan, S. Nakano, N. Miura, H. Ishikuro, and T. Kuroda, "A 12gb/s non-contact interface with coupled transmission lines," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pp. 492 –494, feb. 2011.

[67] T. Matsubara, I. Hayashi, A. Johari, T. Kuroda, and H. Ishikuro, "A 0.7v 4.1mw 850mbps/ch inductive-coupling transceiver with adaptive pulse width controller in 65nm CMOS," in *Radio and Wireless Symposium (RWS), 2012 IEEE*, pp. 71 –74, jan. 2012.

[68] K. Kiyoyama, Y. Tanaka, M. Onoda, T. Fukushima, T. Tanaka, and M. Koyanagi, "A closed-loop power control function for bio-implantable devices," in *Solid-State Circuits Conference, 2008. A-SSCC '08. IEEE Asian*, pp. 325 –328, nov. 2008.

[69] C. Sauer, M. Stanacevic, G. Cauwenberghs, and N. Thakor, "Power harvesting and telemetry in CMOS for implanted devices," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 52, pp. 2605 – 2613, dec. 2005.

[70] M. Zargham and P. Gulak, "Integrated CMOS wireless power transfer for neural implants," in *Biomedical Circuits and Systems Conference (BioCAS), 2011 IEEE*, pp. 165 –168, nov. 2011.

[71] Y. Shimazaki, N. Miura, and T. Kuroda, "A 5.184gbps/ch through-chip interface and automated place-and-route design methodology for 3-d integration of 45nm CMOS processors," in *Cool Chips XV (COOL Chips), 2012 IEEE*, pp. 1 –3, april 2012.

[72] J. Coulombe, M. Sawan, and J.-F. Gervais, "A highly flexible system for microstimulation of the visual cortex: Design and implementation," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 1, pp. 258 –269, dec. 2007.

[73] T. J. Harpster, B. Stark, and K. Najafi, "A passive wireless integrated humidity sensor," *Sensors and Actuators A: Physical*, vol. 95, no. 23, pp. 100 – 107, 2002. Papers from the Proceedings of the 14th IEEE Internat. Conf. on MicroElectroMechanical Systems.

[74] A. Radecki, H. Chung, Y. Yoshida, N. Miura, T. Shidei, H. Ishikuro, and T. Kuroda, "6w/25mm2 inductive power transfer for non-contact wafer-level testing," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pp. 230 –232, feb. 2011.

[75] H. McDermott, "An advanced multiple channel cochlear implant," *Biomedical Engineering, IEEE Transactions on*, vol. 36, no. 7, pp. 789–797, 1989.

[76] M. Clements, K. Vichienchom, W. Liu, C. Hughes, E. McGucken, C. Demarco, J. Mueller, M. Humayun, E. De Juan, J. Weiland, and R. Greenberg, "An implantable power and data receiver and neuro-stimulus chip for a retinal prosthesis system," in *Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on*, vol. 1, pp. 194–197 vol.1, 1999.

[77] K.-Y. Kim, J. Lee, D.-K. Yu, and Y.-S. Park, "Parameter estimation of noisy passive telemetry sensor system using unscented kalman filter," in *Future Generation Communication and Networking (FGCN 2007)*, vol. 2, pp. 433 –438, dec. 2007.

[78] K. Yang, W.-Y. Yin, J. Shi, K. Kang, J.-F. Mao, and Y. Zhang, "A study of on-chip stacked multiloop spiral inductors," *Electron Devices, IEEE Transactions on*, vol. 55, pp. 3236 –3245, nov. 2008.

[79] W.-Y. Yin, J.-Y. Xie, K. Kang, J. Shi, J.-F. Mao, and X.-W. Sun, "Vertical topologies of miniature multispiral stacked inductors," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 56, pp. 475 –486, feb. 2008.

[80] K. Niitsu, Y. Kohama, Y. Sugimori, K. Kasuga, K. Osada, N. Irie, H. Ishikuro, and T. Kuroda, "Modeling and experimental verification of misalignment tolerance in inductive-coupling inter-chip link for low-power 3-d system integration," *Very Large Scale Integration (*VLSI*) Systems, IEEE Transactions on*, vol. 18, pp. 1238 –1243, aug. 2010.

[81] M. Ahmadian, B. Flynn, A. Murray, and D. Cumming, "Data transmission for implantable microsystems using magnetic coupling," *Communications, IEE Proceedings-*, vol. 152, pp. 247 – 250, april 2005.

[82] T. Kuroda, "Wireless proximity communications for 3d system integration," in *Radio-Frequency Integration Technology, 2007. RFIT 007. IEEE International Workshop on*, pp. 21 –25, dec. 2007.

[83] B. Moore, C. Sellathamby, S. Slupsky, and K. Iniewski, "Chip to chip communications for terabit transmission rates," in *Circuits and Systems, 2008. APCCAS 2008. IEEE Asia Pacific Conference on*, pp. 1558 –1561, 30 2008-dec. 3 2008.

[84] A. Chow, D. Hopkins, R. Drost, and R. Ho, "Enabling technologies for multi-chip integration using proximity communication," in VLSI *Design, Automation and Test, 2009.* VLSI-*DAT '09. International Symposium on*, pp. 39 –42, april 2009.

[85] H. Ishikuro and T. Kuroda, "Wireless proximity interfaces with a pulse-based inductive coupling technique," *Communications Magazine, IEEE*, vol. 48, pp. 192 –199, october 2010.

[86] D. Salzman and T. Knight, "Capacitively coupled multichip modules," in *Multichip Modules, 1994. Proceedings of the 1994 International Conference on*, pp. 487 –494, apr 1994.

[87] S. Mick, J. Wilson, and P. Franzon, "4 gbps high-density ac coupled interconnection," in *Custom Integrated Circuits Conference, 2002. Proceedings of the IEEE 2002*, pp. 133 – 140, 2002.

[88] E. Culurciello and A. G. Andreou, "Capacitive inter-chip data and power transfer for 3-d VLSI," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 53, pp. 1348 –1352, dec. 2006.

[89] R. Cardu, M. Scandiuzzo, S. Cani, L. Perugini, E. Franchi, R. Canegallo, and R. Guerrieri, "Chip-to-chip communication based on capacitive coupling," in *3D System Integration, 2009. 3DIC 2009. IEEE International Conference on*, pp. 1 –6, sept. 2009.

[90] G.-S. Kim, M. Takamiya, and T. Sakurai, "A 25-mV-sensitivity 2-Gb/s optimum-logic-threshold capacitive-coupling receiver for wireless wafer probing systems," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 56, pp. 709 –713, sept. 2009.

[91] G.-S. Kim, M. Takamiya, and T. Sakurai, "A capacitive coupling interface with high sensitivity for wireless wafer testing," in *3D System Integration, 2009. 3DIC 2009. IEEE International Conference on*, pp. 1 –5, sept. 2009.

[92] A. Sodagar and P. Amiri, "Capacitive coupling for power and data telemetry to implantable biomedical microsystems," in *Neural Engineering, 2009. NER '09. 4th International IEEE/EMBS Conference on*, pp. 411 –414, 29 2009-may 2 2009.

[93] W. Dubbelday, L. Flesner, G. Garcia, G. Imthurm, and R. Hirschi, "Very high voltage SOS photocell arrays," in *SOI Conference, 1991. Proceedings, 1991., IEEE International*, pp. 84 –85, oct 1991.

[94] M. Marwick and A. Andreou, "Photo-battery fabricated in silicon on sapphire CMOS," *Electronics Letters*, vol. 44, pp. 766 –767, 5 2008.

[95] Y. Arima and M. Ehara, "On-chip solar battery structure for CMOS LSI.," *IEICE electronics express.*, vol. 3, no. 13, pp. 287–291, 2006.

[96] M. Ferri, D. Pinna, E. Dallago, and P. Malcovati, "A 0.35-μm CMOS solar energy scavenger with power storage management system.," *IEEE*, pp. 88–91, 2009.

[97] K. Murari, R. Etienne-Cummings, N. Thakor, and G. Cauwenberghs, "Which photodiode to use: A comparison of CMOS-compatible structures," *Sensors Journal, IEEE*, vol. 9, pp. 752 –760, july 2009.

[98] M. Law and A. Bermak, "High-voltage generation with stacked photodiodes in standard CMOS process," *Electron Device Letters, IEEE*, vol. 31, pp. 1425 –1427, dec. 2010.

[99] F. Horiguchi, "Integration of series-connected on-chip solar battery in a triple-well CMOS LSI," *Electron Devices, IEEE Transactions on*, vol. 59, pp. 1580 –1584, june 2012.

[100] N. Guilar, T. Kleeburg, A. Chen, D. yankelevich, and R. Amirtharajah, "Integrated solar energy harvesting and storage.," *IEEE transaction on very large scale integration (*VLSI*) systems.*, vol. 17, no. 5, pp. 627–637, 2009.

[101] J. Madrenas, D. Fernandez, and C. Wang, "LCMOS: Light-powered standard CMOS circuits," in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, pp. 3029 –3032, may 2012.

[102] A. Fish, S. Hamami, and O. Yadid-Pecht, "CMOS image sensors with self-powered generation capability," *IEEE transactions on circuits and systems*, vol. 53, no. 11, pp. 1210–1214, 2006.

[103] S. Iyer and Y. Xie, "Light emission from silicon," *Science*, vol. 260, no. 5104, pp. 40–46, 1993.

[104] K. Hirschman, L. Tsybeskov, S. Duttagupta, and P. Fauchet, "Silicon-based visible light-emitting devices integrated into microelectronic circuits," 1996.

[105] Z. Lu, D. Lockwood, and J. Baribeau, "Quantum confinement and light emission in sio2/si superlattices," 1995.

[106] A. Richter, P. Steiner, F. Kozlowski, and W. Lang, "Current-induced light emission from a porous silicon device," *Electron Device Letters, IEEE*, vol. 12, pp. 691 –692, dec. 1991.

[107] B. Zheng, J. Michel, F. Y. G. Ren, L. C. Kimerling, D. C. Jacobson, and J. M. Poate, "Room-temperature sharp line electroluminescence at lambda =1.54 microns from an erbium-doped, silicon light emitting diode," *Applied Physics Letters*, vol. 64, pp. 2842 –2844, may 1994.

[108] F. Priolo, G. Franzo, S. Coffa, and A. Carnera, "Erbium implantation in silicon: from materials properties to light emitting devices.," *Elsevier materials chemistry and physics*, vol. 54, pp. 273–279, 1998.

[109] A. Emelyanov, N. Sobolev, and A. Yakimenko, "Capacitance-voltage characteristics of pn structures based on (111) si doped with erbium and oxygen," *Semiconductors*, vol. 35, no. 3, pp. 316–320, 2001.

[110] E. O. Sveinbjornsson and J. Weber, "Room temperature electroluminescence from dislocation-rich silicon," *Applied Physics Letters*, vol. 69, pp. 2686 –2688, oct 1996.

[111] W. Ng, M. Lourenco, R. Gwilliam, S. Ledain, G. Shao, and K. Homewood, "An efficient room-temperature silicon-based light-emitting diode," *Nature*, vol. 410, no. 6825, pp. 192–194, 2001.

[112] Y. Gao, S. Wong, C. W, G. Shao, and K. Homewood, "Effect of implantation temperature on dislocation loop formation and origin of 1.55-m photoluminescence from ion-beam-synthesized fesi2 precipitates in silicon.," *Applied physics letter*, vol. 83, no. 1, pp. 42–44, 2003.

[113] M. Milosavljevi, M. Loureno, G. Shao, R. Gwilliam, and K. Homewood, "Optimising dislocation-engineered silicon light-emitting diodes," *Applied Physics B: Lasers and Optics*, vol. 83, pp. 289–294, 2006. 10.1007/s00340-006-2149-6.

[114] M. Lourenco and G. Shao, "On the role of disocation loops in silicon light emitting diodes.," *Applied physics letters.*, vol. 87, 2005.

[115] K. Homewood and M. Lourenco, "Light from si via dislocation loops.," *Materials today*, no. January, pp. 34–39, 2005.

[116] B. van Drieenhuizen and R. Wolffenbuttel, "Optocoupler based on the avalanche light emission in silicon," *Sensors and Actuators A: Physical*, vol. 31, no. 13, pp. 229 – 240, 1992.

[117] L. W. Snyman, M. Plessis, E. Seevinck, and H. Aharoni, "An efficient low voltage, high frequency silicon CMOS light emitting device and electro-optical interface," *IEEE Electron Device*, vol. 20, no. 12, pp. 614–617, 1999.

[118] M. du Plessis, H. Aharoni, and L. W. Snyman, "A silicon transconductance light emitting device (transled)," *Sensors and Actuators A: Physical*, vol. 80, no. 3, pp. 242 – 248, 2000.

[119] M. du Plessis, H. Aharoni, and L. W. Snyman, "Two- and multi-terminal CMOS/biCMOS si leds," *Optical Materials*, vol. 27, no. 5, pp. 1059 – 1063, 2005. Si-based Photonics: Towards True Monolithic Integration Proceedings of the European Materials Research Society Symposium A1 European Materials Research Society 2004 Spring Meeting.

[120] H. Aharoni and M. du Plessis, "Low-operating-voltage integrated silicon light-emitting devices," *Quantum Electronics, IEEE Journal of*, vol. 40, no. 5, pp. 557–563, 2004.

[121] M. Green, J. Zhao, A. Wang, P. Reece, and M. Gal, "Efficient silicon light-emitting diodes.," *Nature*, vol. 412, pp. 805–808, 2001.

[122] H. Rong, Y. H. Kuo, S. Xu, A. Liu, R. Jones, and M. Paniccia, "Monolithic integrated raman silicon laser," *OPTICS EXPRESS*, vol. 14, no. 15, pp. 6705–6712, 2006.

[123] L. Pavesi, "Silicon-based light sources for silicon integrated circuits," *Advances in Optical Technologies*, vol. 2008, 2008.

[124] D. Schroder, N. Thomas, and J. Swartz, "Free carrier absorption in silicon.," *IEEE journal on solid state circuits*, vol. SC13, pp. 180–187, 1978.

[125] R. A. Soref and B. R. Bennett, "Electrooptical effects in silicon," *IEEE journal of quantum electronics*, vol. QE-23, no. 1, pp. 123–129, 1987.

[126] J. Isenberg and W. Warta, "Free carrier absorption in heavily doped silicon layers.," *Applied physics letters.*, vol. 84, 2004.

[127] J. P. Lorenzo and R. A. Soref, "1.3 μm electro-optic silicon switch," *Applied Physics Letters*, vol. 51, pp. 6 –8, jul 1987.

[128] Q. Xu, B. Schmidt, S. Pradhan, and M. Lipson, "Micrometer-scale silicon electro-optic modulator," *Nature letters*, vol. 435, pp. 325–327, 2005.

[129] G. Reed, G. Mashanovich, F. Gardes, and D. Thomson, "Silicon optical modulators," *Nature Photonics*, vol. 4, no. 8, pp. 518–526, 2010.

[130] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson, "12.5 gbit/s carrier-injection-based silicon micro-ring silicon modulators," *Opt. Express*, vol. 15, no. 2, pp. 430–436, 2007.

[131] S. Manipatruni, Q. Xu, B. Schmidt, J. Shakya, and M. Lipson, "High speed carrier injection 18 gb/s silicon micro-ring electro-optic modulator," in *Lasers and Electro-Optics Society, 2007. LEOS 2007. The 20th Annual Meeting of the IEEE*, pp. 537 –538, oct. 2007.

[132] A. Cutollo, M. Iodice, P. Spirito, and L. Zeni, "Silicon electro-optic modulator based on a three terminal device integrated in a low-loss single-mode soi waveguide," *Journal of lightwave technology*, vol. 15, pp. 505–518, 1997.

[133] A. Sciuto, S. Libertino, A. Alessandria, S. Coffa, and G. Coppola, "Design, fabrication, and testing of anintegrated si-based light modulator," *Journal of lightwave technology*, vol. 21, no. 1, p. 228, 2003.

[134] F. Gardes, G. Reed, N. Emerson, and C. Png, "A sub-micron depletion-type photonic modulator in silicon on insulator," *Opt. Express*, vol. 13, pp. 8845–8854, Oct 2005.

[135] L. Liao, A. Liu, D. Rubin, J. Basak, Y. Chetrit, H. Nguyen, R. Cohen, N. Izhaky, and M. Paniccia, "40 gbit/s silicon optical modulator for high-speed applications," *Electronics letters*, vol. 43, no. 22, 2007.

[136] D. Thomson, F. Gardes, J.-M. Fedeli, S. Zlatanovic, Y. Hu, B. Kuo, E. Myslivets, N. Alic, S. Radic, G. Mashanovich, and G. Reed, "50-gb/s silicon optical modulator," *Photonics Technology Letters, IEEE*, vol. 24, pp. 234 –236, feb.15, 2012.

[137] D. Marris-Morini, L. Vivien, J. M. Fédéli, E. Cassan, P. Lyan, and S. Laval, "Low loss and high speed silicon optical modulator based on a lateral carrier depletion structure," *Opt. Express*, vol. 16, pp. 334–339, Jan 2008.

[138] N.-N. Feng, S. Liao, D. Feng, P. Dong, D. Zheng, H. Liang, R. Shafiiha, G. Li, J. E. Cunningham, A. V. Krishnamoorthy, and M. Asghari, "High speed carrier-depletion modulators with 1.4v-cm vπl integrated on 0.25μm silicon-on-insulator waveguides," *Opt. Express*, vol. 18, pp. 7994–7999, Apr 2010.

[139] P. Dong, S. Liao, D. Feng, H. Liang, D. Zheng, R. Shafiiha, C.-C. Kung, W. Qian, G. Li, X. Zheng, A. V. Krishnamoorthy, and M. Asghari, "Low vpp, ultralow-energy, compact, high-speed silicon electro-optic modulator," *Opt. Express*, vol. 17, pp. 22484–22490, Dec 2009.

[140] G. Rasigade, D. Marris-Morini, L. Vivien, and E. Cassan, "Performance evolutions of carrier depletion silicon optical modulators: From p-n to p-i-p-i-n diodes," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 16, pp. 179 –184, jan.-feb. 2010.

[141] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, "A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor," *Nature*, vol. 427, no. 6975, pp. 615–618, 2004.

[142] J. Fujikata, J. Ushida, Y. Ming-Bin, Z. ShiYang, D. Liang, P. Guo-Qiang, D.-L. Kwong, and T. Nakamura, "25 ghz operation of silicon optical modulator with projection mos structure," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pp. 1 –3, march 2010.

[143] K. Kajikawa, T. Tabei, and H. Sunami, "An infrared silicon optical modulator of metal–oxide–semiconductor capacitor based on accumulation-carrier absorption," *Japanese Journal of Applied Physics*, vol. 48, no. 4, p. 04C107, 2009.

[144] S. Fathpour and B. Jalali, "Energy harvesting in silicon optical modulators," *Opt. Express*, vol. 14, pp. 10795–10799, 2006.

[145] R. A. Soref, "Silicon-based optoelectronics," *Proceedings of the IEEE*, vol. 81, no. 12, pp. 1687–1706, 1993.

[146] G. Reed, F. Gardes, Y. Hu, D. Thomson, L. Lever, R. Kelsall, and Z. Ikonic, "Silicon photonics: optical modulators," in *Proceedings of SPIE*, vol. 7608, p. 76080J, 2010.

[147] W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators," *Laser & Photonics Reviews*, vol. 6, no. 1, pp. 47–73, 2012.

[148] R. Baker, CMOS *circuit design, layout, and simulation.* Hoboken, John Wiley and Sons, Inc., 3 ed., 2011.

[149] G. Palumbo and D. Pappalardo, "Charge pump circuits: An overview on design strategies and topologies," *Circuits and Systems Magazine, IEEE*, vol. 10, pp. 31 –45, quarter 2010.

[150] J. Dickson, "On-chip high-voltage generation in mnos integrated circuits using an improved voltage multiplier technique," *Solid-State Circuits, IEEE Journal of*, vol. 11, pp. 374 – 378, jun 1976.

[151] T. Tanzawa and T. Tanaka, "A dynamic analysis of the dickson charge pump circuit," *Solid-State Circuits, IEEE Journal of*, vol. 32, pp. 1231 –1240, aug 1997.

[152] T. Tanzawa, "Dickson charge pump circuit design with parasitic resistance in power lines," in *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, pp. 1763 –1766, may 2009.

[153] T. Tanzawa, "A switch-resistance-aware dickson charge pump model for optimizing clock frequency," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 58, pp. 336 –340, june 2011.

[154] M. Makowski, "Realizability conditions and bounds on synthesis of switched-capacitor dc-dc voltage multiplier circuits," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 44, pp. 684 –691, aug 1997.

[155] L. Liu and Z. Chen, "Analysis and design of makowski charge-pump cell," in *ASIC, 2005. ASICON 2005. 6th International Conference On*, vol. 1, pp. 497 –502, oct. 2005.

[156] A. Umezawa, S. Atsumi, M. Kuriyama, H. Banba, K. Imamiya, K. Naruke, S. Yamada, E. Obi, M. Oshikiri, T. Suzuki, and S. Tanaka, "A 5-v-only operation 0.6- mu;m flash eeprom with row decoder scheme in triple-well structure," *Solid-State Circuits, IEEE Journal of*, vol. 27, pp. 1540 –1546, nov 1992.

[157] S. Atsumi, M. Kuriyama, A. Umezawa, H. Banba, K. Naruke, S. Yamada, Y. Ohshima, M. Oshikiri, Y. Hiura, T. Yamane, *et al.*, "A 16-mb flash eeprom with a new self-data-refresh scheme for a sector erase operation," *IEICE transactions on electronics*, vol. 77, no. 5, pp. 791–799, 1994.

[158] R. Pelliconi, D. Iezzi, A. Baroni, M. Pasotti, and P. Rolandi, "Power efficient charge pump in deep submicron standard CMOS technology," *Solid-State Circuits, IEEE Journal of*, vol. 38, pp. 1068 – 1071, june 2003.

[159] R. Pelliconi, "Us patent 6,819,162." patent, September 2004.

[160] J. Che, C. Zhang, Z. Liu, Z. Wang, and Z. Wang, "Ultra-low-voltage low-power charge pump for solar energy harvesting systems.," *IEEE*, 2009.

[161] H. Morimura and N. Shibata, "A step-down boosted-wordline scheme for 1-v battery-operated fast sram's," *Solid-State Circuits, IEEE Journal of*, vol. 33, pp. 1220 –1227, aug 1998.

[162] J. Starzyk, Y.-W. Jan, and F. Qiu, "A dc–dc charge pump design based on voltage doublers.," *IEEE transactions on circuits and systems.*, vol. 48, no. 3, pp. 350–359, 2001.

[163] Y. Ramadass and A. Chandrakasan, "Voltage scalable switched capacitor dc-dc converter for ultra-low-power on-chip applications," in *Power Electronics Specialists Conference, 2007. PESC 2007. IEEE*, pp. 2353 –2359, june 2007.

[164] Y. Nakagome, H. Tanaka, K. Takeuchi, E. Kume, Y. Watanabe, T. Kaga, Y. Kawamoto, F. Murai, R. Izawa, D. Hisamoto, T. Kisu, T. Nishida, E. Takeda, and B. Kiyoo Itoh, "An experimental 1.5-v 64-mb dram," *Solid-State Circuits, IEEE Journal of*, vol. 26, no. 4, pp. 465–472, 1991.

[165] P. Favrat, P. Deval, and M. Declercq, "A high-efficiency CMOS voltage doubler," *Solid-State Circuits, IEEE Journal of*, vol. 33, no. 3, pp. 410–416, 1998.

[166] H.-M. Lee and M. Ghovanloo, "An adaptive reconfigurable active voltage doubler/rectifier for extended-range inductive power transmission," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 59, no. 8, pp. 481–485, 2012.

[167] H. Lee and P. Mok, "Switching noise and shoot-through current reduction techniques for switched-capacitor voltage doubler," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 5, pp. 1136–1146, 2005.

[168] T. Esram and P. Chapman, "Comparison of photovoltaic array maximum power point tracking techniques," *Energy Conversion, IEEE Transactions on*, vol. 22, pp. 439 –449, june 2007.

[169] V. Salas, E. Olas, A. Barrado, and A. Lzaro, "Review of the maximum power point tracking algorithms for stand-alone photovoltaic systems," *Solar Energy Materials and Solar Cells*, vol. 90, no. 11, pp. 1555 – 1578, 2006.

[170] W. Xiao, A. Elnosh, V. Khadkikar, and H. Zeineldin, "Overview of maximum power point tracking technologies for photovoltaic power systems," in *IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society*, pp. 3900 –3905, nov. 2011.

[171] N. S. D'Souza, L. A. Lopes, and X. Liu, "Comparative study of variable size perturbation and observation maximum power point trackers for pv systems," *Electric Power Systems Research*, vol. 80, no. 3, pp. 296 – 305, 2010.

[172] H. Shao, C.-Y. Tsui, and W.-H. Ki, "The design of a micro power management system for applications using photovoltaic cells with the maximum output power control," *IEEE transactions on very large integration (*VLSI*) systems.*, vol. 17, no. 8, pp. 1138–1142, 2009.

[173] M. Pastre, F. Krummenacher, O. Kazanc, N. Pour, C. Pace, S. Rigert, and M. Kayal, "A solar battery charger with maximum power point tracking," in *Electronics, Circuits and Systems (ICECS), 2011 18th IEEE International Conference on*, pp. 394 –397, dec. 2011.

[174] M. Forbes, J. Gourlay, and M. Desmulliez, "Optically interconnected electronic chips: a tutorial and review of the technology," *Electronics Communication Engineering Journal*, vol. 13, pp. 221 –232, oct 2001.

[175] C. Kromer, G. Sialm, C. Berger, T. Morf, M. Schmatz, F. Ellinger, D. Erni, G.-L. Bona, and H. Jackel, "A 100-mw 4 times;10 gb/s transceiver in 80-nm CMOS for high-density optical interconnects," *Solid-State Circuits, IEEE Journal of*, vol. 40, pp. 2667 – 2679, dec. 2005.

[176] R. Bockstaele, M. De Wilde, W. Meeus, O. Rits, H. Lambrecht, J. Van Campenhout, J. De Baets, P. Van Daele, E. van den Berg, M. Klemenc, *et al.*, "A scalable parallel optical interconnect family," *IO overview paper*, pp. 1–10, 2004.

[177] P. Ellinghaus, P. Venter, M. Du Plessis, P. Rademeyer, A. Bogalecki, *et al.*, "A fully CMOS optical transmission system based on light emitting avalance diodes," *SAIEE Africa Research Journal*, vol. 101, pp. 17–20, March 2010.

[178] S. Tam and M.-C. Chang, "Rf/wireless-interconnect: The next wave of connectivity," *SCIENCE CHINA Information Sciences*, vol. 54, pp. 1026–1038, 2011. 10.1007/s11432-011-4225-8.

Chapter 3

Theoretical background: physics and CMOS manufacturing

This chapter will lay out the fundamental theoretical background that underlies the entire PhD thesis. This will begin with a study of the absorption of electro-magnetic waves in semiconductors whereby the phenomena utilised within the context of this thesis, namely inter-band absorption of light in Silicon and free carrier absorption will be presented. Both phenomena are derived from basic physical principles and key equations are visited. Next, the theory of semiconductor junctions will be tackled as junction-based devices will be used to manipulate how EM waves propagate through and interact with the semiconductive material. Both p-n and p-i-n topologies will be studied so much at a basic level with many simplifying assumptions as much as at a higher level with some of the assumptions lifted. Moreover, a brief background of doping in CMOS technologies will be given whereby we shall be using fundamental physics and readily available expressions that can help determine the doping profile of a pn-junction if certain manufacturing process details are given. The ideal processes of planar diffusion, ion implantation and a combination of the two are considered and example doping profiles are plotted for single dopant integration. Finally, the basics of photo-transduction in Silicon will be laid out.

3.1 Absorption of EM waves in semiconductors

As electro-magnetic radiation propagates through a semiconductor it will be subjected to, amongst other phenomena, some degree of amplitude decay. This occurs because semiconductive materials have the ability to absorb incoming EM radiation through a variety of mechanisms. Inter-band absorption and free-carrier absorption specifically will be the phenomena studied in this section. The Franz-Keldysh effect is also mentioned as it too occurs in bulk Silicon, but the quantum confined Stark effect is not considered since it requires a quantum well and is therefore inapplicable in CMOS technology. The theory developed in this section introduces the relevant semiconductor physics and is largely based on key texts, including [1].

3.1.1 Inter-band absorption

It is well known that electrons in semiconductor crystals possess an energy band structure. The highest occupied energy level at zero temperature is the upper edge of the valence band and lowest unoccupied energy level at the zero temperature is the lower edge of the conduction band. A valence electron can jump over the energy gap separating these two bands (denoted by E_g hence forth) and reach the conduction band by absorbing a photon of sufficient energy:

$$E = hf = \frac{hc}{\lambda} \geq E_g \tag{3.1}$$

where E is the energy of the photon, h represents Planck's constant, f is the frequency of the photon c represents the speed of light and λ is its wavelength. This process is inter-band absorption. Note: photons of lower than required energies will not be able to cause excitation unless they can cause transitions from the conduction band to some other allowable 'stray' energy state introduced in the forbidden gap by lattice defects or impurities. Photons whose energy is very high, on the other hand, may completely eject the target electron from the crystal. In this case the extra energy, above what is strictly needed to eject the electron from the atom is transformed into kinetic energy for the electron.

When a photon whose energy is above the bare minimum as dictated by E_g causes an inter-band electron transition, then typically the electron will occupy a higher energy state than the least energetic state of the conduction band, often denoted as E_C (the most energetic state of the valence band is similarly dubbed E_V). In general when this occurs the electron will gradually descend towards the E_C state while losing energy in the form of heat. Figure 3.1 shows allowable energy transitions in a crystal featuring energy bands. Note: transitions such as the so-called 'extrinsic transitions' that occur between a valid energy state and a 'stray state' introduced within the band structure by some crystal defect or impurity are shown in Figure 3.1, but will not be considered in this section. Similarly the ionising transition whereby an electron receives so much energy as to be ejected from the system altogether is also shown in Figure 3.1 but shall not be discussed further in this section.

Within the context of a crystal we can say that the number of photons absorbed by a mono-layer of semiconductive material will be proportional to the optical flux Φ_0 falling upon it. The units of Φ_0 are photons per unit area per second ($\frac{n}{m^2 s}$ in S.I.). For each slice of semiconductor of thickness Δx that incoming radiation passes through the losses will be proportional to the optical flux entering the slice, as well as to the thickness of the slice:

$$\Phi(x + \Delta x) - \Phi(x) = -\alpha \Phi(x) \Delta x \tag{3.2}$$

where $\Phi(x)$ is the optical flux at depth x within the material and α is the proportionality constant linking optical losses to the initial optical flux and the thickness of the slice. The negative sign denotes that the phenomenon is one of optical loss while α, also dubbed as the 'absorption coefficient' is a positive number.

Equation 3.2 gives rise to the following differential form:

$$\frac{d\Phi(x)}{dx} = -\alpha \Phi(x) \tag{3.3}$$

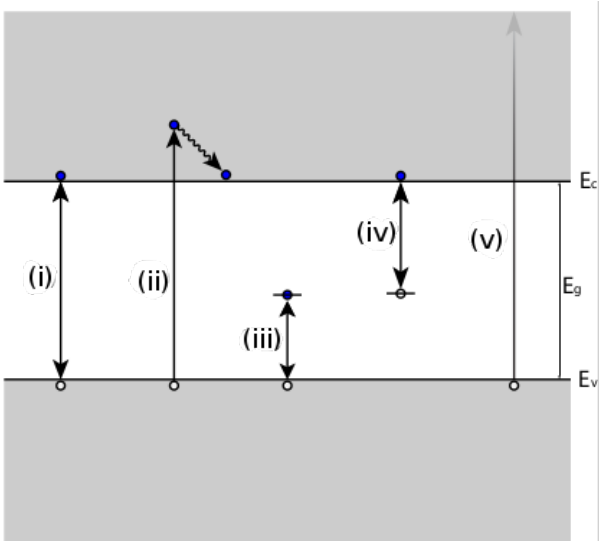

Figure 3.1: Energy band diagram showing the allowable energy state transitions in a semiconductive crystal. Shown are: (i) Bare-minimum energy band-to-band transition. (ii) High energy band-to-band transition followed by thermal relaxation. (iii, iv) Transitions involving 'stray' energy levels introduced by impurities or lattice defects. (v) Ejection from the crystal. Notation: E_g: Band-gap. E_V: Valence band. E_C: Conduction band.

which then leads to the solution in the form of:

$$\boxed{\Phi(x) = \Phi_0 e^{-\alpha x}} \tag{3.4}$$

which is also known as the Beer-Lambert relation. As a boundary condition we used optical flux at a depth of 0 equal to Φ_0. The amount of light exiting a slab of semiconductor can be found using equation 3.4 by setting x = L, where L is the length of the optical path through the slab, or equivalently the thickness of the slab if we send the beam perpendicularly through the material.

The absorption coefficient (α) is strongly dependent on photon energy. Notably, as the photon energy drops below E_g (or equivalently the photon wavelength increases beyond the critical wavelength that corresponds to a photon energy of E_g) α drops very abruptly. Conversely as photon energy increases above E_g absorption occurs more readily (α increases).

3.1.2 Free-carrier absorption

When EM radiation passes through a non-depleted volume of semiconductive material it will suffer amplitude losses even if the photon energies involved are below the energy gap, and thus inter-band transition is not occurring. This is due to the phenomenon of free-carrier absorption. To put this into perspective one may note that the energy gap in Silicon is approximately 1.11 eV at room temperature (25^oC) which corresponds to photon wavelength of approximately 1120nm. Much of the analysis given in this section is based on the work of Soref [2].

To understand free-carrier absorption one may start from the solution to Maxwell's equations for the propagation of a plane EM wave through a conductive medium in the x-direction [3]:

$$\vec{E}(x,t) = \vec{E}_0 \exp\left(i(\omega t - kx)\right) \tag{3.5}$$

where $\vec{E}$ is the electric field, $\vec{E}_0$ is the base amplitude of the electric field, ω is the angular frequency of the EM wave, t represents time, k is the wavenumber of $\vec{E}$ and x is the distance the beam traversed through the medium from the point where it's vector was given by $\vec{E}_0$. i here and henceforth represents the imaginary unit.

The complex refractive index n is defined as:

$$n = \frac{ck}{\omega} \tag{3.6}$$

where c is the speed of light in vacuum. After substituting 3.6 into 3.5 we obtain:

$$\vec{E}(x,t) = \vec{E}_0 \exp\left(-\frac{\omega n_i x}{c}\right) \exp\left[i\omega\left(t - \frac{n_r x}{c}\right)\right] \tag{3.7}$$

where n_i and n_r are the imaginary and real parts of the refractive index. The 1st exponential represents amplitude attenuation. The 2nd exponential represents the 'slowdown' that the EM wave is subject to when crossing the medium, i.e. its phase.

From 3.7, the amplitude of the EM wave at position x is given by:

$$A(x) = |\vec{E}(x,t)| = A_0 \exp\left(-\frac{\omega n_i x}{c}\right) \tag{3.8}$$

where $A(x)$ is the amplitude of the electric field as a function of distance traversed through the material and A_0 is its base amplitude. This amplitude component is then squared to yield the beam intensity:

$$I(x) = A^2(x) = A_0^2 \exp\left(-2\frac{\omega n_i x}{c}\right) = I_0 \exp\left(-2\frac{\omega n_i}{c}x\right) = I_0 \exp(-\alpha x) \tag{3.9}$$

where $I(x)$ is the intensity of the beam as a function of penetration depth and the newly appeared α variable is the coefficient of absorption. Explicitly:

$$\alpha = \frac{2\omega n_i}{c} \tag{3.10}$$

Equation 3.9 is of the familiar Beer-Lambert form.

Revisiting Maxwell's equations for the propagation of EM waves in conductive media where the conductivity is a complex value σ we can find out that the complex refractive index takes the (implicit) form:

$$n^2 = \frac{\epsilon\mu}{c^2}\left(1 + i\frac{\sigma}{\epsilon\omega}\right) \tag{3.11}$$

where ϵ represents the electrical permittivity of the material under study, μ is the magnetic permeability of the said material and σ is the conductivity of the material. n^2 can be decomposed into its real and imaginary components:

$$n^2 = (n_r + n_i)^2 \approx n_r^2 + i2n_r n_i \tag{3.12}$$

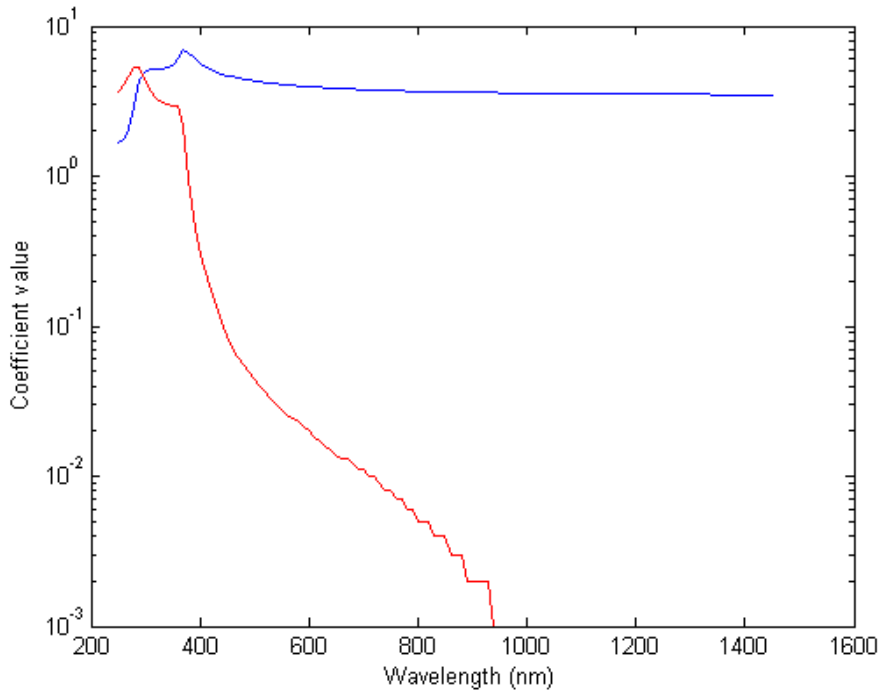

Figure 3.2: The refractive index of Silicon at 300^oK shown broken apart into real and imaginary parts. The blue trace represents the real part of the refractive index (refraction) whilst the red trace represents the imaginary part (absorption). Note how the imaginary part of the index has completely tailed off by approximately 1000nm wavelength. No data for higher wavelengths is specified in [4] because the absorption effect becomes vanishingly small. The steps towards 1000nm wavelength are due to quantisation errors.

if we assume that $|n_i| \ll |n_r|$. This is not an unreasonable assumption at the wavelengths of interest ($\geq$1120nm) as can be seen from the tables that appear in [4]. For the convenience of the reader the tabulated data has been plotted in Figure 3.2.

By combining 3.12 and 3.11 with 3.10 and using the substitutions $\epsilon = \epsilon_r\epsilon_0$, $\mu = \mu_r\mu_0$ and $\epsilon\mu/c = \epsilon_r$ (assuming $\mu_r \approx 1$) we obtain an expression for α that is connected to the conductivity of the material:

$$\alpha = \frac{Re(\sigma)}{\epsilon_0 nc} \tag{3.13}$$

If we assume that the Drude model for an AC field of frequency ω holds, the conductivity can also be expressed as:

$$\sigma(\omega) = \frac{n_e e^2 \tau_e}{m_e(1 - i\omega\tau)} \tag{3.14}$$

where n_e is the free electron concentration, e is the charge of a single electron, τ_e is the relaxation time (average time between two electron collisions) and m_e is the electron's effective mass given by:

$$m_e = \frac{e\tau_e}{\mu_e} \Leftrightarrow \mu_e = \frac{e\tau_e}{m_e} \tag{3.15}$$

where μ_e is the mobility of free electrons within the material. By combining 3.14 and 3.15 into 3.13, using $\omega = 2\pi c/\lambda$, assuming $\omega\tau \gg 1$ and putting the equation in difference form with respect to free electron concentration we obtain:

$$\Delta\alpha = \left(\frac{e^3\lambda^2}{4\pi^2 c^3 \epsilon_0 n}\right)\left[\frac{\Delta n_e}{m_e^2 \mu_e}\right] \tag{3.16}$$

which expresses how the absorption coefficient in Silicon changes when the concentration of free-carriers is manipulated. The equation's non-difference version appears in a publication by Schroeder [5]. An extension to the formula so that both free electrons and free holes are taken into account has been presented in a key publication by Soref and Bennet [6]:

$$\Delta\alpha = \left(\frac{e^3\lambda^2}{4\pi^2 c^3 \epsilon_0 n}\right)\left[\frac{\Delta n_e}{m_e^2 \mu_e} + \frac{\Delta n_h}{m_h^2 \mu_h}\right] \tag{3.17}$$

where n_h, m_h and μ_h represent the concentration, effective mass and mobility of holes.

This equation leads to certain conclusions: The square dependence on wavelength, the inversely proportional dependence on carrier mobility and the linear dependence on carrier concentration reveal opportunities for engineering the phenomenon either at the manufacturing level or at the design level.

The mobility can only be engineered at manufacturing level. Carrier concentration is determined by the manufacturing of the semiconductor, but can also be controlled within very narrow volumes of semiconductor skirting the edges of pn-junctions by use of appropriate biasing. Specifically changing the magnitude of the reverse bias across a pn-junction we can modulate the width of its depletion region and thus cause material that was previously undepleted and full of free carriers to become depleted (or the other way around). Finally the choice of operating wavelength presents some opportunity for engineering after fabrication. The only restriction imposed by the material is that the photon energies involved are below that of the band-gap but high enough for the purposes of the specific implementation (e.g. km-order, radio-wave wavelengths are unlikely to lend themselves for use in applications concerning integrated circuits). Note: free-carrier absorption continues to occur at photon energies above the band-gap, but inter-band absorption then becomes the dominant phenomenon, thus somewhat reducing its potential for use in practical applications.

3.1.3 Franz-Keldysh effect

Bulk Silicon exhibits one more important electro-optical effect: the Franz-Keldysh effect. This effect causes Silicon to absorb photons of energy below the nominal band-gap more readily if a strong electric field is applied to it. As detailed in Keldysh's publication [7] electric fields in the region of $10^7\, V/m$ are required before the effect becomes of any significance. The intensity of the electric field is linked to the band-edge shift, i.e. the practical cut-off point beyond which photons carry too little energy to cause inter-band absorption. This phenomenon is of little effect within the context of this thesis as it wil tend not to affect the dominance of free-carrier absorption at the wavelengths of interest ($\lambda \geq 1300\, nm$ - safely above the $\approx 1100\, nm$ wavelength at which inter-band absorption phenomena become significant).

3.2 Theory of semiconductor junctions

The interface between a p-type region and an n-type region residing within the same semiconductive crystal is called a 'junction'. If one were to trace a route between a point in bulk p-type to a point in bulk[1] n-type the concentration of p-dopants would start from a high equilibrium value and then drop as the junction is approached. At some point the concentration of p-dopants N_A will be equal to the concentration of n-dopants N_D. The surface defined by all points where $N_A = N_D$ holds is called the 'metallurgical surface'. Beyond the metallurgical point the concentration of p-dopants typically continues to drop until it stabilises at its n-type bulk concentration. The concentration of n-dopants generally follows the same qualitative changes, but in the opposite direction. The absolute difference in doping concentrations $|N_A - N_D|$ yields the net doping concentration. We can plot doping concentrations vs. location along a line connecting p-bulk to n-bulk that crosses the metallurgical surface at a right angle to obtain $N_D(x)$ for n-type (a.k.a. 'donor') dopants, and $N_A(x)$ for p-type (a.k.a. 'acceptor') dopants. On the basis of the shapes of $N_D(x)$ and $N_A(x)$ we can categorise junctions in two important classes: Abrupt and non-abrupt (otherwise known as 'graded') junctions. In abrupt junctions changes in doping concentrations happen in steps whilst in graded junctions such restriction does not apply.

Another significant differentiation between semiconductor junctions concerns whether between the distinctly p- and n-type regions there exists a layer of mostly intrinsic material, where the net doping concentration is very close to zero. If such layer doesn't exist, then the junction is of the traditional p-n type. If such layer does exist, then the condition $N_A = N_D$ is satisfied over a volume of material rather than over a surface. This volume is called the 'intrinsic region' of the junction which itself is then dubbed a p-i-n junction for that reason.

The analysis developed in this section introduces the fundamental background to semiconductor junctions based on [1].

3.2.1 Junctions at thermal equilibrium

P- and n-type materials feature different concentrations of free carriers. For that reason in every semiconductor junction regardless of type there will be diffusion of free carriers across the junction. This gives rise to the diffusion currents (a free electron and a free hole diffusion current to be precise). However, this exchange of charged particles creates an electric field across the junction, which gives rise to the electron and hole drift currents. These drift currents tend to counteract the diffusion currents, following essentially a variant of the Le Chatelier principle. At thermal equilibrium and if any external influences are discounted the net current across any junction must be equal to zero. Both the electron and the hole components of the junction currents need to cancel out individually as electron and hole concentrations either side of the junction eventually settle at equilibrium values. If we normalise currents by the cross-sectional area of the conducting medium we obtain the following equations that relate to electron and hole current densities at equilibrium:

$$J_e = J_{e,diff} + J_{e,drift} = q\mu_e n_e E - qD_e \frac{dn_e}{dx} \tag{3.18}$$

[1] By bulk here a region of purely unipolarly doped semiconductor (e.g. no counterdoping) as would result from the connection of two pieces of semiconductor that have been manufactured independently.

$$J_h = J_{h,diff} + J_{h,drift} = q\mu_h n_h E - qD_h \frac{dn_h}{dx} \tag{3.19}$$

where J_e and J_h are the electron and hole current densities, subsequently split into diffusion and drift components ($J_{x,diff}$ and $J_{x,drift}$), q is the fundamental unit of charge, μ_x represents mobility for holes or electrons (electrons if $x =$ 'e', holes if $x =$ 'h'), n_x represents carrier concentrations, E is the electric field intensity and D_x is the diffusion constant.

The diffusion coefficient of free carriers follows the Einstein relation:

$$D_x = \frac{kT\mu_x}{q} \tag{3.20}$$

where k is Boltzmann's constant and T is the absolute temperature. Substituting Einstein's relation into 3.18 and 3.19 we obtain:

$$J_x = q\mu_x n_x E - kT\mu_x \frac{dn_x}{dx} \tag{3.21}$$

When an electric field is applied to a piece of semiconductor, free carriers are subjected to an electrostatic force. By definition this force is equal to the negative gradient of the carrier's potential energy. For an electron at the edge of the conduction band (potential energy = E_C by definition) this is described by the following equation:

$$-qE = -\frac{dE_C}{dx} \tag{3.22}$$

which implies directly that the electric field takes the form:

$$E = \frac{1}{q}\frac{dE_C}{dx} \tag{3.23}$$

However, in a material that features a constant or nearly constant band-gap (an assumption which holds for the doping concentrations normally seen in integrated Silicon semiconductors, which are no higher than about $10^{22}\, dopants/cm^3$) the gap between the edge of the conduction band and the intrinsic Fermi energy remains constant, which means that:

$$\frac{dE_C}{dx} \approx \frac{dE_i}{dx} \tag{3.24}$$

where E_i is the intrinsic Fermi energy level. The Fermi energy (E_F) is the theoretical energy level which, if available, would feature a 50% probability of being occupied by an electron according to the Boltzmann distribution. In intrinsic semiconductors E_F is denoted as E_i. By introducing 3.24 into 3.23 we obtain:

$$E = \frac{1}{q}\frac{dE_i}{dx} \tag{3.25}$$

Furthermore, the well-known equations describing free carrier concentration as a function of E_i and E_F are:

$$n_e = n_i \exp\left(\frac{E_F - E_i}{kT}\right) \tag{3.26}$$

for free electrons and:

$$n_h = n_i \exp\left(\frac{E_i - E_F}{kT}\right) \tag{3.27}$$

for free holes, where n_i denotes the common free electron and free hole concentration in an intrinsic semiconductor at temperature T. For free electrons the derivative $\frac{dn_e}{dx}$ is given by:

$$\frac{dn_e}{dx} = \frac{n_e}{kT}\left(\frac{dE_F}{dx} - \frac{dE_i}{dx}\right) \tag{3.28}$$

If we substitute 3.28 and 3.25 into 3.21 taken for free electrons we obtain:

$$J_e = \mu_e n_e \frac{dE_F}{dx} \tag{3.29}$$

and for holes we can obtain a similar result by going through the same calculations:

$$J_h = \mu_h n_h \frac{dE_F}{dx} \tag{3.30}$$

In equilibrium both J_e and J_h need to equal 0 which leads to the condition of thermal equilibrium:

$$\boxed{\frac{dE_F}{dx} = 0} \tag{3.31}$$

This condition fundamentally determines the nature of the so-called 'depletion region', a region of semiconductor in the proximity of a pn-junction where there are very few free carriers.

The electric field within the semiconductor is described by Poisson's equation:

$$-\frac{dE(x)}{dx} = -\frac{\rho_s(x)}{\epsilon} = -\frac{q}{\epsilon}(N_D(x) - N_A(x) + n_h(x) - n_e(x)) \tag{3.32}$$

where ρ_s is the net charge density at the point where the gradient of the electric field is taken. The term in parentheses in the rightmost branch of the equation above represents the net density of charged elements. Assuming that the 'full ionisation' condition holds, i.e. that all dopant atoms are ionised the N_x terms represent charge contributed by immobile atomic cores while the n_x terms represent the contributions made by mobile carriers.

In bulk-n sections of the semiconductor, we can make the approximation that $N_A \approx 0$ and $n_e \gg n_h$. Also charge neutrality holds. This implies:

$$n_e = N_D \tag{3.33}$$

in n-type bulk and similarly:

$$n_h = N_A \tag{3.34}$$

in p-type bulk. Equations 3.33 and 3.34 are essentially re-statements of the full ionisation condition for pure n- and pure p-type materials.

If we set the rightmost side of equation 3.18 equal to zero and substitute 3.28 we obtain:

$$\boxed{E = \frac{1}{q}\left(\frac{dE_i}{dx} - \frac{dE_F}{dx}\right)} \tag{3.35}$$

which shows that the difference between the rates of change of the intrinsic Fermi energy level and the actual Fermi energy level is proportional to the electric field. From that we can integrate both sides, sign-invert and set the integration constant to zero in order to obtain the electrostatic potential difference between an intrinsic piece of semiconductor and a piece where the Fermi energy level is equal to E_F:

$$\psi_{N_D,i} = \frac{1}{q}(E_F - E_i) \tag{3.36}$$

where $\psi_{N_D,i}$ is the electrostatic potential difference between an n-type semiconductor with doping concentration N_D and an intrinsic semiconductor. The sign inversion occurs because $\psi_{N_D,i}$ is by definition equal to the negative gradient of the electric field $-dE/dx$. By using 3.26 and 3.33 into 3.36 we obtain:

$$\psi_{N_D,i} = \frac{kT}{q} ln\left(\frac{N_D}{n_i}\right) \tag{3.37}$$

Similar calculations for p-bulk Silicon yield:

$$\psi_{i,N_A} = \frac{1}{q}(E_i - E_F) = \frac{kT}{q} ln\left(\frac{N_A}{n_i}\right) \tag{3.38}$$

By then combining equations 3.36, 3.37 and 3.38 we can calculate the electrostatic potential difference between p-bulk and n-bulk that will appear across a semiconductive junction:

$$\boxed{V_{bi} = \psi_{N_D,i} + \psi_{i,N_A} = -\frac{1}{q}(E_{Fn} - E_{Fp}) = \frac{kT}{q} ln\left(\frac{N_A N_D}{n_i^2}\right)} \tag{3.39}$$

where we have marked the n- and p-side Fermi levels as E_{Fn} and E_{Fp} respectively.

This potential difference is called the 'built-in potential' and the leftmost part of the equation above shows it in its well-known form. Notably, the built in potential can be defined between any pair of points in a semiconductor irrespective of what doping concentrations or how many discernible pn-junctions lie in between. Most often, however, it refers to the built-in potential between a pair of points lying in bulk Silicon of opposite polarities across a single p-n or p-i-n junction.

Crucially, the boxed equations 3.31, 3.35 and 3.39 hold for any type of junction regardless of doping profile.

The depletion region

Lying between p-bulk and n-bulk material there will be a region, centred around the metallurgical surface that is largely depleted of mobile carriers. This phenomenon arises because carriers that have diffused across the metallurgical surface are captured and binded by immobile donor or acceptor atoms. The density of this trapped charge is very close to the doping concentration because under normal circumstances the vast majority of donor or acceptor atoms bind a mobile carrier. This implies that the net charge density within the depletion region can be expressed as:

$$\rho_s(x) = q(N_D(x) - N_A(x)) \tag{3.40}$$

where ρ_s is a function of location, takes positive values in n-type material (where the dopant atoms bind holes) and negative values in p-type material (where dopant atoms bind electrons). In typical semiconductors, however, the doping concentration either side of a junction is so heavily skewed in favour of one type of dopant over the other that we can simplify equation 3.40:

$$\rho_s(x) = qN_D(x) \tag{3.41}$$

for the n-type side and:

$$\rho_s(x) = qN_A(x) \tag{3.42}$$

for the p-type side. For the entire depletion region Poisson's equation (3.32) becomes (this time including both electric field and electrostatic potential forms):

$$-\frac{d\psi^2_{(x,ref)}}{dx^2} = \frac{dE(x)}{dx} = -\frac{q}{\epsilon}(N_A(x) - N_D(x)) \tag{3.43}$$

where $\psi_{(x,ref)}$ is the difference in electrostatic potential between point 'x' and point 'ref' (often this will be abberviated to $\psi(x)$). In this case point 'ref' is an arbitrary reference point. This is different from ψ_{N_x,N_y} -without parentheses- where the subscript variables denote doping concentrations rather than locations.

Crucially, the depletion region is subject to the 'overall charge neutrality' condition. We know that bulk material of any type must be overall charge-neutral (same number of electrons and protons in the crystal). When creating a pn-junction no charge is created or destroyed and therefore the overall neutrality of the fused materials is preserved. These statements directly imply that the depletion region must also observe overall charge neutrality. Given this constraint and equations 3.41 and 3.42 we obtain the following expression:

$$\boxed{\int_{-x_p}^{0} q|N_A(x)|dx = \int_{0}^{x_n} q|N_D(x)|dx} \tag{3.44}$$

where we have assumed the usual trajectory from a point in p-bulk to a point in n-bulk, 0 is the origin of the coordinate system, negative values reside in p-type material, positive values reside in n-type material and x_p and x_n denote the widths of the depletion region's p- and n-sides respectively.

Note: the depletion region (a.k.a. space charge region) has 'smooth edges', that is to say the transition from depleted semiconductor to bulk does not happen abruptly, but rather gives rise to a so-called 'transition region'. Nevertheless, the transition region in typical Silicon junctions is very narrow compared to the depletion region and for that reason depletion regions are almost invariably treated using a 'sharp edge approximation' whereby the transition region is considered to be of zero width.

Finally, it must be remarked that the term 'transition region' is somewhat subjective as there will be a continuous distribution of free carriers of both types throughout the entire depletion region and even extending into bulk material of the opposite polarity. Nevertheless, under zero or reverse bias conditions by the time one moves away from p-bulk (or n-bulk) material by a distance equivalent to the transition region width the concentration of free holes has already dropped to a negligible amount.

Semiconductor junction vital information summary

In summary we can extract much vital information from any semiconductor junction by following this process, shown in Figure 3.3 in the form of a flow-chart:

1. Begin by setting up a geometry of the junction under study. Typically this is done by defining a trajectory from a point in p-bulk to a point in n-bulk which crosses the metallurgical surface perpendicularly, then defining the intersection point as the origin and finally setting that for $x < 0$ the material is p-type and for $x > 0$ the material is n-type.
2. Define the doping concentration functions $N_A(x)$ and $N_D(x)$ (align them to the frame of reference). From these we can directly compute the built-in potential between any pair of points in the bulk $v_{bi}(x_1, x_2)$ (equation 3.39).
3. Parametrically define a depletion region by setting x_n and x_p. We can then generate the bulk free-carrier distribution function equal to the sum of $N_A(x)$ and $N_D(x)$ for $x \notin [-x_p, x_n]$ and the depletion region net charge density function $\rho_s(x) \in [-x_p, x_n]$, which equals $N_D(x) - N_A(x)$ so that the polarity of the net charge density is taken into account. The variables x_n and x_p are linked together via the charge neutrality relation (equation 3.44), so defining one sets the other one too.
4. Integrate $\rho_s(x)$ with respect to x and obtain the electric field profile $E(x)$ and then integrate $-E(x)$ with respect to x to obtain the electrostatic potential distribution $\psi_{(x,ref)}$ where ref is a reference point.
5. Apply the boundary condition that $\psi_{(-x_p,x_n)} = V_{bi}(-x_p, x_n)$ for $x_p \neq n_n$ and determine the actual width of the depletion region W.

Thus we obtain the following vital junction information: a) The built-in potential as a function of location $V_{bi}(x_1, x_2)$. b) The charge density in the depletion region $\rho_x(x)$. c) The electric field distribution $E(x)$. d) The electrostatic potential variation $\psi_{(x,ref)}$. e) The width of the depletion region W.

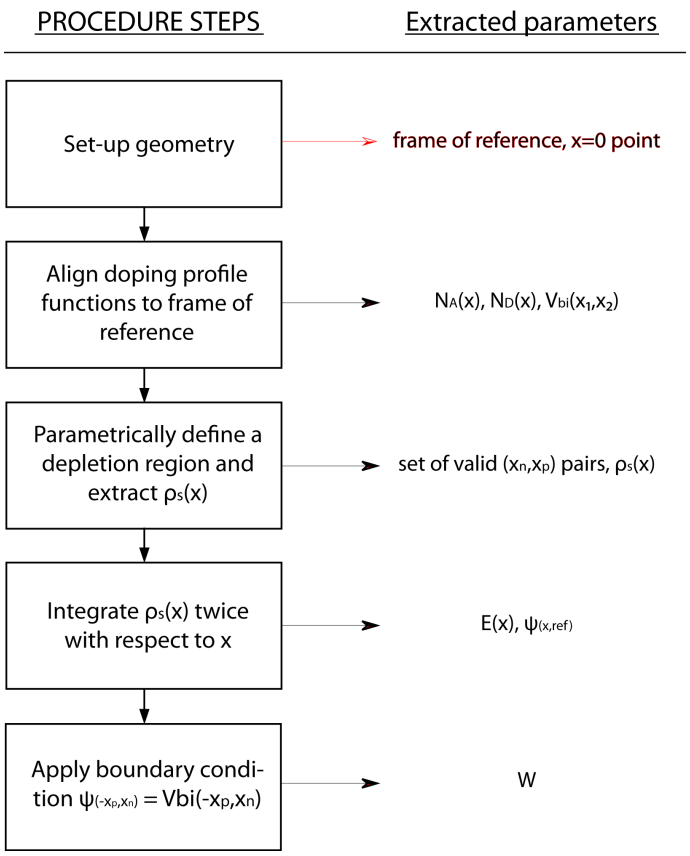

Figure 3.3: Process of determining the vital statistics of a semiconductive junction starting from the doping profile.

3.2.2 Ideal (abrupt) p-n and p-i-n junctions

As was mentioned before, abrupt junctions are those where doping concentrations as a function of location change in steps. Typical doping profiles of abrupt p-n and p-i-n type junctions can be seen in Figure 3.4 whilst the resulting energy band diagram can be seen in Figure 3.5.

Figure 3.4 shows an abrupt pn-junction in physical and band-diagram form. In (a) the disks with the '+' or '-' signs inside them denote donor and acceptor atoms respectively, small white dots represent free holes and small black dots represent free electrons. The lack of free carriers in the depletion regions is shown explicitly, but an electron-hole pair within the depletion region is still shown in order to illustrate the direction of the electrostatic forces that act upon them. In (b) E_{Cp} and E_{Cn} denote the edges of the conduction band in p- and n-type material. E_{Vp} and E_{Vn} similarly denote valence band edges in p- and n-type regions. The effect of the electric field present in the depletion region on the conduction (and valence) band edges can be seen in the form of non-zero dE_C/dx (equation 3.35). E_F denotes the Fermi energy level, which in accordance to equation 3.31 remains constant throughout the entire length of the material. E_g is the band-gap. $q\psi_{bi}$ is the built-in potential difference multiplied by the charge, a quantity that denotes the difference in energy level between an electron residing at corresponding energy levels in bulk-p and bulk-n material. The arrow at the bottom left indicates the x-direction. The dashed line running vertically through the middle of both panels represents the metallurgical surface. Adapted from [8]. Note that the concentration of dopants of either type never falls below a minimum 'background concentration' of contaminant atoms. The graphs stress this in order to indicate that abrupt junctions need not necessarily be between pure p- and pure n-type materials. Similarly the intrinsic region of a pin-junction need not be perfectly intrinsic to play the role of intrinsic region. In the pin-junction graph a distinctive 'valley' between regions of high concentration of p- and n-type dopants marks the intrinsic region.

Essentially an abrupt pn-junction can be described as a slab of completely homogeneous p-type material coming into contact with a corresponding slab of n-type material. Therefore, dopant concentrations N_A and N_D become just a function of the side they are situated in. As such, the Poisson equation becomes:

$$\frac{dE}{dx} = -\frac{qN_A}{\epsilon} \tag{3.45}$$

in the p-side and:

$$\frac{dE}{dx} = \frac{qN_D}{\epsilon} \tag{3.46}$$

in the n-side. The combination of these equations proves the fact that the electric field within an abrupt pn-junction takes a triangular form since dE/dx is constant within each side of the depletion region (and zero in the bulk). If we define a trajectory from bulk-p to bulk-n that intersects the metallurgical surface perpendicularly, and then we define the intersection point as $x = 0$, p-type material as any point with $x < 0$ and n-type material any point at $x > 0$ we can solve 3.45 and 3.46 to obtain the electric field profile along the trajectory:

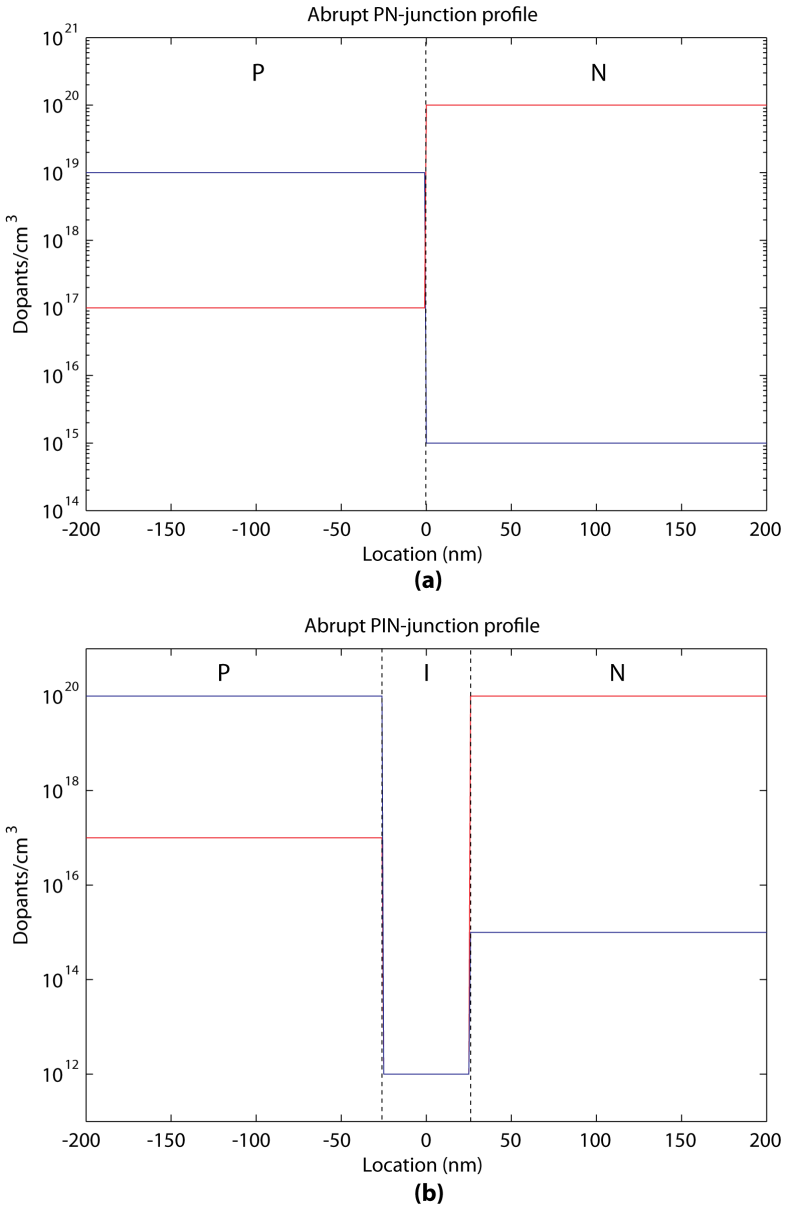

Figure 3.4: Acceptor and donor doping concentrations as a function of location in an abrupt pn-junction (a) and an abrupt pin-junction (b). In this case the blue lines denote p-type dopant atoms (acceptors) whilst the red lies denote n-type dopants (donors).

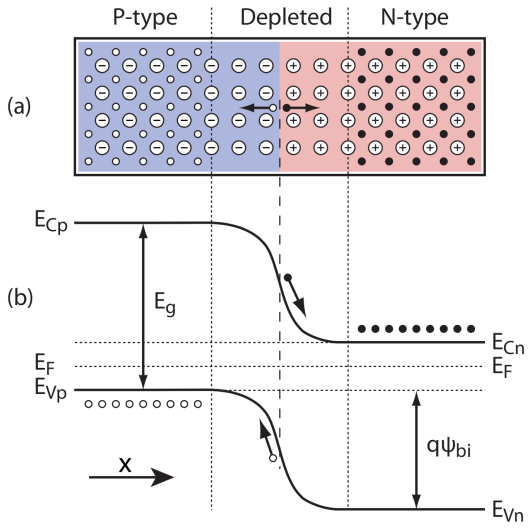

Figure 3.5: The abrupt pn-junction shown in (a) its physical form and (b) its corresponding band diagram.

$$E(x) = -\frac{qN_D}{\epsilon}(x_n - x) \mid x \in [0, x_n] \tag{3.47}$$

for the n-side where x_n is the width of the n-side of the depletion region. For the p-side:

$$E(x) = -\frac{qN_A}{\epsilon}(x_p + x) \mid x \in [-x_p, 0] \tag{3.48}$$

where x_p is the width of the depletion region in the p-side.

Equations 3.47 and 3.48 yield the following expressions for the maximum magnitude of the electric field within the junction, found at the metallurgical surface. Since the electric field profile must be a continuous function of location (otherwise $d\psi_{(x,0)}(x)/dx$ would have infinite magnitude) this maximum value must be equal for both computation by use of p-side and by use of n-side equations, yielding a boundary condition:

$$\boxed{E_{max} = \frac{qN_D x_n}{\epsilon} = \frac{qN_A x_p}{\epsilon}} \tag{3.49}$$

The electric field profile within an abrupt pn-junction as dictated by equations 3.47 and 3.48 is shown in Figure 3.6, which also shows the physical manifestation of an abrupt pn-junction along with its free carrier concentration, net charge density and electrostatic potential as functions of location.

The 'overall charge neutrality' condition (equation 3.44) for abrupt junctions takes the form:

$$N_A x_p = N_D x_n \tag{3.50}$$

The total depletion region width is the sum of the widths of the p- and n-side depletion regions:

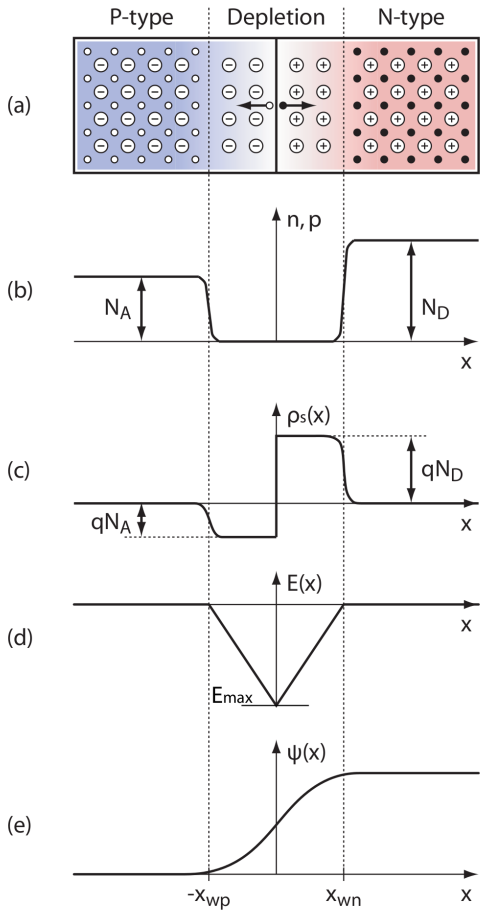

Figure 3.6: The pn-junction. Shown are: (a) the physical manifestation, (b) net free carrier concentration vs location, (c) net charge density vs location, (d) electric field density vs location and (e) electrostatic potential vs location. x_{wn} and x_{wp} are the edges of the depletion region on the n- and p-sides respectively. Plots are qualitative only. The smoothness of the curves illustrates the existence of a transition region. Adapted from [8].

$$W = x_n + x_p \tag{3.51}$$

By combining equations 3.47 and 3.48 into a single expression for electric field as a function of location, integrating appropriately and using 3.51 we obtain an expression for the built-in potential:

$$V_{bi} = -\int_{x_n}^{x_p} E(x)dx = -\int_{-x_p}^{0} E(x)dx - \int_{0}^{x_n} E(x)dx = \frac{qN_A x_p^2}{2\epsilon} + \frac{qN_D x_n^2}{2\epsilon} = \frac{E_{max}W}{2} \tag{3.52}$$

Combining 3.52 with 3.50 then yields the well-known equation for the zero external bias depletion region width of an abrupt pn-junction:

$$\boxed{W = \sqrt{\frac{2\epsilon}{q}\left(\frac{N_A + N_D}{N_A N_D}\right) V_{bi}}} \tag{3.53}$$

This equation tells us that the width of the depletion width depends on the doping concentrations either side of the metallurgical surface and the voltage across the terminals of the pn-junction. If we add external bias to the system equation 3.53 becomes:

$$\boxed{W = \sqrt{\frac{2\epsilon}{q}\left(\frac{N_A + N_D}{N_A N_D}\right) (V_{bi} - V_{ext})}} \tag{3.54}$$

where V_{ext} is the applied external bias. The sign is negative if we assume the external bias voltage variable to represent forward bias, in which case the external voltage will work to counteract the built-in potential.

Special case: single sided pn-junctions

Very often the case is that one side of the junction is much more heavily doped than the other side. For example in cases where N_A is over 1 order of magnitude greater than N_D equation 3.53 simplifies to:

$$\boxed{W \approx x_n \approx \sqrt{\frac{2\epsilon V_{bi}}{qN_D}}} \tag{3.55}$$

Indicating that the heavily doped side makes a minor contribution to the width of the depletion region. On the other hand, the electric field remains in triangular form, but now the steepness of the curve is much greater on the narrow, heavily doped side than on the wide, lightly doped side. This means that the integral of the electric field, and therefore also it's negative, the electrostatic potential, is primarily contributed by the lightly doped side, in other words, most of the potential drop across the pn-junction occurs across the lightly doped region.

Given the one-sided nature of the junction, the maximum electric field can be approximated as:

$$\boxed{E_{max} = \frac{qN_D x_n}{\epsilon} \approx \frac{qN_D W}{\epsilon}} \tag{3.56}$$

If we now once again create a trajectory that connects a point in p-bulk to a point in n-bulk while intersecting the metallurgical surface perpendicularly with the intersection point defined as the origin and the lightly doped n-side occupying the positive 1/2-axis we can express the electric field as a function of location:

$$E(x) = \frac{qN_D}{\epsilon}(x - x_n) \approx \frac{qN_D}{\epsilon}(x - W) = -E_{max}\left(1 - \frac{x}{W}\right) \tag{3.57}$$

which can then be integrated and sign-inverted to yield a good approximation of the electrostatic potential as a function of location:

$$\psi_{(x,0)} = \int_0^x E dx = E_{max}\left(x - \frac{x^2}{W}\right) \tag{3.58}$$

using the assumption that practically all of the electrostatic potential drop occurs across the lightly-doped region we may approximate:

$$\psi_{(-x_p,0)} \approx 0 \Leftrightarrow \psi_{(W,0)} \approx V_{bi} \tag{3.59}$$

Using 3.52 to replace E_{max} we finally obtain:

$$\boxed{\psi_{(0,x)} = \frac{V_{bi}x}{W}\left(2 - \frac{x}{W}\right)} \tag{3.60}$$

Finally, adding external biasing to the system we obtain the equivalent of equation 3.54 for single-sided junctions:

$$W = \sqrt{\frac{2\epsilon}{qN_D}(V_{bi} - V_{ext})} \tag{3.61}$$

Similar results can be obtained for single-sided junctions where the lightly doped region is p-type.

p-i-n structures

By adding a slice of intrinsic material between a p-type and an n-type slab we obtain a junction that features a region of genuinely intrinsic material (as opposed to merely charge neutral). Diffusion currents will still try to flow across it from the p- and n-type regions that surround it, carriers will eventually diffuse across the intrinsic region and start being bound by the dopant atoms of the opposite side, an electric field will result from this movement, a drift current will be set into motion by this electric field and finally a depletion region in the p- and n-type regions will form. As such, the basic physical phenomena behind pin-junctions are the same as for pn-structures. The only difference is that between depleted p- and depleted n-type material there will now be a slice of intrinsic semiconductor. This intrinsic slice,

in contrast to depleted Silicon, will feature no fixed charge in its domain of the depletion region. The implication is immediately visible if we consider:

$$\frac{dE}{dx} = \frac{q}{\epsilon} \sum_{x \in p} z_x N_x \tag{3.62}$$

In this more generalised form of equations 3.46 and 3.45 p is the set of all dopant atoms and N_x is a dopant x of valence z_x. The valence parameter is an integer that takes positive values for donor atoms and negative values for acceptor atoms and corresponds to the net charge left if all mobile carriers associated with the said dopant atom are taken away (as would be the case in depleted material) in units of q. The connection to the intrinsic region of a p-i-n junction is that the right hand side of the equation is by definition zero, which yields the condition that within the intrinsic region of a p-i-n junction the electric field density stays constant. Therefore the electric field profile in a p-i-n junction is of a trapezoidal form.

Once again, in order to prove this it is useful to set-up a trajectory from p-bulk to n-bulk that intersects the surfaces that denote transition from doped to undoped Silicon perpendicularly. Notably: unlike in pn-junctions, there is no longer a unique metallurgical surface, but rather an intrinsic volume featuring a p-i and an i-n delimiting surface. The depletion region can therefore be now split into 3 segments:

For $x \in [x_p, 0]$:

$$\frac{dE}{dx} = -\frac{qN_A}{\epsilon} \tag{3.63}$$

For $x \in [0, x_i]$, where x_i is the width of the intrinsic region:

$$\frac{dE}{dx} = 0 \tag{3.64}$$

And for $x \in [x_i, x_i + x_n]$:

$$\frac{dE}{dx} = -\frac{qN_D}{\epsilon} \tag{3.65}$$

where the charge neutrality condition holds exactly as in the abrupt pn case:

$$N_A x_p = N_D x_n \tag{3.66}$$

but this time the depletion region width also has to take the intrinsic part into consideration:

$$W = x_n + x_i + x_p \tag{3.67}$$

The maximum electric field density relation remains unchanged from the pn-case (equation 3.49), but now the built-in potential is expressed with an additional term in the integral:

$$V_{bi} = -\int_{-xp}^{0} E(x)dx - \int_{0}^{x_i} E_{max}dx - \int_{x_i}^{x_i+x_n} E(x)dx = \\ = \frac{qN_a x_p^2}{2\epsilon} + E_{max}x_i + \frac{qN_D x_n^2}{2\epsilon} = E_{max}\left(\frac{x_p + 2x_i + x_n}{2}\right) \tag{3.68}$$

If we assume that the intrinsic region is the dominant component, then the we obtain:

$$W \approx x_i \tag{3.69}$$

and:

$$V_{bi} \approx E_{max}W \tag{3.70}$$

Notably, so long as the contributions of the p- and n-side depletion regions are considered to remain negligibly small, the width of the depletion region becomes independent of the built-in voltage and also independent of any external bias voltage. Moreover, the electrostatic potential is then described by the following approximation (for the boundary condition $\psi_{(0,ref)} = 0$ where ref denotes a reference point):

$$\boxed{\psi_{(x,0)} = \frac{x}{W}V_{bi}} \tag{3.71}$$

for $x \in [0, W]$ which under external biasing becomes:

$$\boxed{\psi_{(x,0)} = \frac{x}{W}(V_{bi} - V_{ext})} \tag{3.72}$$

The electrostatic potential changes linearly through the pin-junction both under external biasing and when left unbiased.

Finally, the energy band diagram and much of the vital information of a sample pin-junction are shown in Figures 3.7 and 3.8.

3.2.3 Significance of doping profiles for free carrier absorption

The doping concentration profile of any volume of semiconductor is a crucial determinant of its behaviour. Amongst other things electro-optical modulation capabilities are influenced. Here we seek to give simple examples of the qualitative nature of these interactions. We shall study abrupt, homogeneous junctions with planar metallurgical surfaces and draw conclusions on the nature of doping concentrations -particularly around the depletion region- that are amenable for good modulation performance.

Electro-optical modulation hinges upon the application of two different bias voltage across a pn-junction. These voltages will correspond to unique depletion region widths across our test pn-junction and thus also unique optical losses when passing through the semiconductor volume.

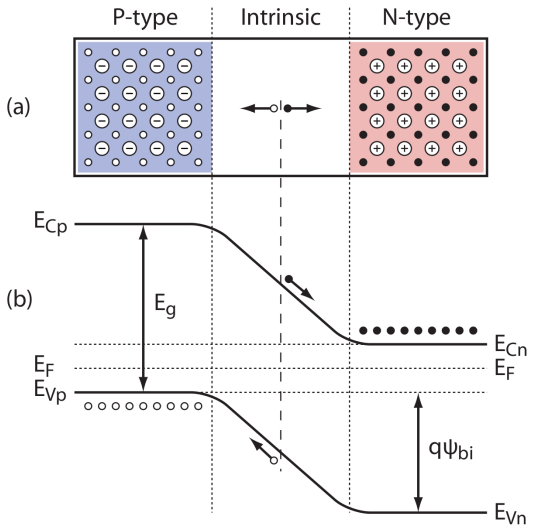

Figure 3.7: The pin-junction. Shown are: (a) the physical manifestation and (b) the energy band diagram. Note the fact that the band slope remains unchanged throughout the intrinsic region, in contrast to Figure 3.5 that shows the shape of the energy bands in the depleted region of a pn-junction. Adapted from [8].

We begin by setting up a simple system where the p-side and the n-side widths are equal to D, the doping concentration of both the p-side and the n-side is N_C (and therefore the width of the depletion region is the same either side of the junction). We will be biasing the device at V_1 ('ON') and V_2 ('OFF') where $V_1 < V_2$. Light crosses the junction perpendicular to the metallurgical surface.

Given the above parameters we can use equations 3.54 and 3.39 in order to express the depletion region width as:

$$W = \sqrt{\frac{2\epsilon}{q}\left(\frac{2}{N_C}\right)\left(\frac{kT}{q}ln\left(\frac{N_C^2}{n_i^2}\right) - V_{bias}\right)} = \sqrt{\frac{4\epsilon}{qN_C}\left(2V_T ln\left(\frac{N_C}{n_i}\right) - V_{bias}\right)} \tag{3.73}$$

where V_T is the thermal voltage and V_{bias} will be V_1 or V_2.

From this we can derive that:

$$w = W(V_1) - W(V_2) = \sqrt{\frac{A}{N_C}}\left(\sqrt{V_{bi} - V_1} - \sqrt{V_{bi} - V_2}\right) \tag{3.74}$$

which yields a convenient expression for the difference in optically resistive path lengths for our two bias conditions.

The transmittance through the material will be given by:

$$T(V_{bias}) = e^{-a_{h+}\left(\frac{D-W(V_{bias})}{2}\right) - a_{e-}\left(\frac{D-W(V_{bias})}{2}\right)} = e^{-\frac{a_{h+}+a_{e-}}{2}(D-W(V_{bias}))} = e^{-a(D-W(V_{bias}))} \tag{3.75}$$

with a_{h+} and $a_{e=}$ being the absorption coefficients for hole and electrons respectively at doping concentration N_C and $a = \frac{a_{holes}+a_{electrons}}{2}$.

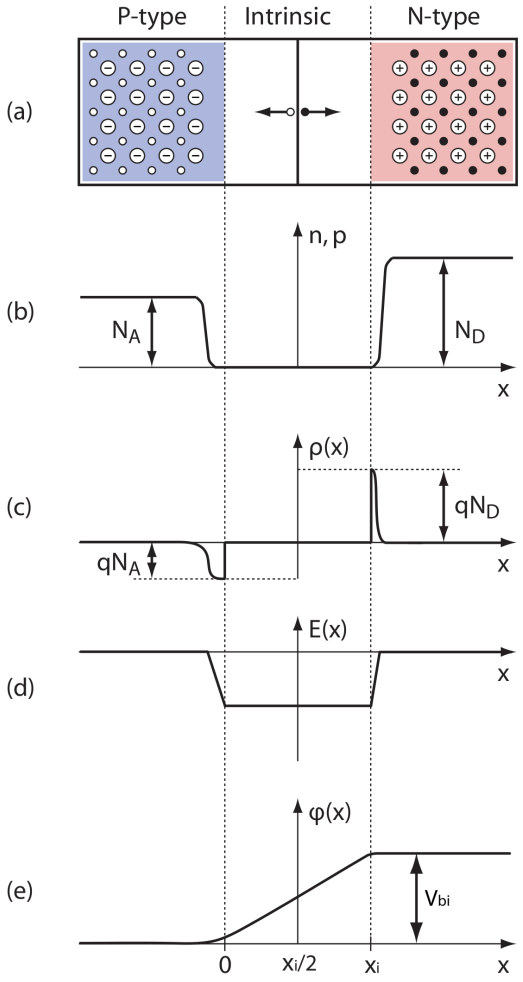

Figure 3.8: The pin-junction: (a) Physical manifestation and (b-e) important graphs describing a pin-junction. We observe: (b) the mobile carrier concentration, (c) net charge density, (d) the electric field profile and (e) the electrostatic potential, all as functions of location. This image is best compared with the pn-junction equivalent in Figure 3.6. Note the drastic change in the charge density vs. location function (c), the trapezoidal shape of E(x) with its asymmetry around $x_i/2$ due to doping concentration differences between p- and n-sides and the largely linear nature of $\psi(x)$ Adapted from [8].

Therefore, the difference in transmittance between the two bias points (absolute modulation depth) can expressed as:

$$\Delta T = e^{-a(D-W(V_1))} - e^{-a(D-W(V_2))} = e^{-a(D-W(V_2))}\left(e^{aw}-1\right) \approx e^{-a(D-W(V_2))}(aw) \tag{3.76}$$

where w is the difference $W(V_1) - W(V_2) > 0$ given $V_1 < V_2$. Under the assumption that w is small (which it will be in CMOS -see worked example in chapter 4, section 4.2 for realistic values of a for CMOS-order doping concentrations and W under various bias conditions), we can approximate that $e^{aw} \approx 1 + aw$.

If we then normalise ΔT by the transmittance under high bias, i.e. $T(V_2) = e^{-a(D-W(V_2))}$ we obtain the fractional modulation depth:

$$\boxed{\frac{\Delta T}{T} \approx aw} \tag{3.77}$$

Knowing from 4.1 on page 107 that the absorption coefficient will be roughly proportional to doping concentration and using 3.74, ignoring the dependence of V_{bi} on $ln(N_C)$ we can obtain the approximate result that:

$$\frac{\Delta T}{T} \propto \frac{N_C}{\sqrt{N_C}} = \sqrt{N_C} \tag{3.78}$$

Thus we have proved that so long as all our approximations hold true, higher doping concentrations in abrupt pn-junctions will mean greater fractional modulation depth for the same bias conditions. This result can be extrapolated into the remark that the fractional modulation depth will increase if the doping concentrations at the volume swept by the edge surfaces of the depletion region when the modulator is swept between some bias voltages V_1 and V_2 (the 'sweep volume') are higher[2]. This remark still assumes that the doping concentration in the sweep volume is roughly unchanged, but it is intuitive to think that any increase in doping concentration within a given sweep volume will lead to increased modulation. Further study is required, however, in order to reveal whether this extrapolation is true and if and how it generalises to sweep volumes of arbitrary doping concentrations and sizes.

Note: because the built-in potential increases with doping concentration in reality the depletion region width difference w decreases more slowly than with the square root of the doping concentration. For that reason we actually expect better that square-root dependence of fractional modulation depth on doping concentration.

Similar analysis can be performed on non-symmetric, abrupt, homogeneous pn-junctions to yield the following result:

$$\boxed{\frac{\Delta T}{T} \approx a_p w_p + a_n w_n} \tag{3.79}$$

[2]And not necessarily equal as in the simple case used for our proof.

where a_x is the absorption coefficient on the x side of the junction (p-side: $x = p$, n-side: $x = n$) and w_x is the depletion region width change between high and low bias on the x side of the junction.

Using the charge neutrality equation 3.44 on page 71 for the abrupt, homogeneous case at high and low bias conditions and subtracting we can easily prove that the following holds:

$$w_p N_A = w_n N_D \Leftrightarrow \frac{w_p}{w_n} = \frac{N_D}{N_A} \tag{3.80}$$

which intuitively states that the volumes of semiconductor swept by the depletion region edge between high and low bias voltages either side of the junction must contain the same amount of space charge. Crucially, we also observe that the ratio of w_p/w_n is always locked to N_D/N_A. Additionally we have by definition:

$$w = w_p + w_n \tag{3.81}$$

i.e. the sum of the p-side and n-side depletion region width changes between high and low bias conditions equals the total change of depletion region width over the entire pn structure.

Combining equation 3.80 with 3.81 we can easily prove that:

$$w_p = \frac{w N_D}{N_A + N_D} \tag{3.82}$$

If once again we ignore the dependence of V_{bi} on the doping concentrations either side of the junction we can use equation 3.54 on page 78 to show:

$$w = \sqrt{\frac{2\epsilon}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)}\left(\sqrt{V_{bi} - V_1} - \sqrt{V_{bi} - V_2}\right) \propto \sqrt{\frac{N_A + N_D}{N_A N_D}} \tag{3.83}$$

Combining equations 3.82 and 3.83 we can show that w_p depends on the doping concentrations either side of the pn-junction according to:

$$w_p \propto \sqrt{\frac{N_D/N_A}{N_A + N_D}} \tag{3.84}$$

Finally, if we take into consideration that free carrier absorption coefficients are proportional to free carrier concentration (p-side: $a_p \propto N_A$, n-side $a_n \propto N_D$) we find that the important expression $a_p w_p$ related to doping concentrations according to the following formula[3]:

$$\boxed{a_p w_p \propto \sqrt{\frac{N_A N_D}{N_A + N_D}}} \tag{3.85}$$

[3]The accurate formula, if we do include the variation of V_{bi} on doping concentrations takes the form: $a_p w_p \propto \sqrt{\frac{N_A N_D}{N_A + N_D}}\Delta\left(\sqrt{V_{bi} - V}\right)$ with V representing external biasing.

which yields an expression on how doping concentrations either side of the junction affect the modulation contribution of the p-side of the junction. We can easily that expression for the n-side is exactly the same.

Using the charge neutrality equation version shown in 3.80 and the fact that electron and hole free carrier absorption coefficients are proportional to their respective free carrier concentrations we we can show:

$$\boxed{\frac{a_p w_p}{a_n w_n} = P\frac{N_A}{N_D}\frac{N_D}{N_A} = P} \tag{3.86}$$

where P is the proportionality constant between free carrier absorption coefficients (as can be derived from equation 4.1 on page 107 or from Soref's formula [6]). This is important because it tells us that the ratio of p-side to n-side contributions to electro-optical modulations in an abrupt, homogeneous junction is always equal (at least theoretically) to the ratio of free carrier absorption coefficients for common doping concentration.

3.2.4 Effects of illumination spectral content on measured electro-optical modulation

We shall consider here how the spectral content of the incoming radiation affects measured fractional electro-optical modulation depth in two important cases: a) the illumination is monochromatic and b) the illumination is broadband.

Monochromatic case

What we are measuring at the photodetector is a photocurrent generated by incoming electro-magnetic radiation. This will depend on the initial intensity of the radiation Q_0, the overall transmittance from emitter to photodetector $T(V_{bias})$ and on the sensitivity of the photodetector to the wavelength at which the emitter operates $S(\lambda)$ (we assume a) no beam divergence and b) that all light passes through the modulator for simplicity).

We now seek to find the fractional modulation depth F, which we measure by the formula:

$$F = \frac{\Delta I}{I} = \frac{Q_0 T(V_1) S(\lambda_0) - Q_0 T(V_2) S(\lambda_0)}{Q_0 T(V_2) S(\lambda_0)} = \frac{\Delta T}{T} \tag{3.87}$$

where I is the photocurrent measured and is a function of the bias voltage applied to the modulator. ΔI is the difference in photocurrents between high and low bias conditions. λ_0 is the wavelength of operation. We have denoted the reference transmittance level $T(V_2)$ simply as T in the rightmost part of the expression. For small values of F it makes little difference which state occupies the denominator of the expression.

The key finding is that measured F corresponds to the $\frac{\Delta T}{T}$ expression we used previously in this section to denote fractional modulation depth. Crucially, ΔT is entirely the result of electro-optical modulation, and yet it scales with T^4. T itself depends on steady-state losses along the entire optical

[4] Under the assumption that all light that reaches the photodetector is being modulated.

path. Unfolding the term $\frac{\Delta T}{T}$ we obtain:

$$F(\lambda) = \frac{\Delta T}{T}(\lambda) \approx a_p(\lambda)w_p + a_n(\lambda)w_n \tag{3.88}$$

where it can be easily shown that any 'steady state' losses, common to both bias conditions drop out of the equation.

Knowing that free carrier absorption coefficients for both electrons and holes vary with the square of the wavelength (see equation 4.1) we can take the ratio of fractional electro-optical modulation depths in the same test device for two different wavelengths, λ and x times λ:

$$\boxed{\frac{F(x\lambda)}{F(\lambda)} = \frac{a_p(x\lambda)w_p + a_n(x\lambda)}{a_p(\lambda) + a_n(\lambda)} = \frac{x^2(a_p(\lambda) + a_n(\lambda))}{(a_p(\lambda) + a_n(\lambda))} = x^2} \tag{3.89}$$

where we have shown that altering the operating wavelength simply scales up or down the measured fractional electro-optical modulation depth uniformly for all devices regardless of geometry or composition.

Broadband illumination case

When the illumination is broadband the measured electro-optical modulation expression changes:

$$F = \frac{\Delta I}{I} = \frac{\int_\Lambda Q(\lambda)\Delta T(\lambda)S(\lambda)d\lambda}{\int_\Lambda Q(\lambda)T(\lambda)S(\lambda)d\lambda} \tag{3.90}$$

Due to the integral we can no longer conveniently simplify this expression to $\frac{\Delta T}{T}$, which demonstrates that the specific spectrum of our broadband illumination does affect the measured electro-optical modulation.

3.3 Doping processes in CMOS technology

Every semiconductor fabrication process, be it CMOS, bipolar or BiCMOS includes well-defined procedures for infusing the Silicon substrate with dopant atoms. More complicated processes may include large numbers of highly specialised and optimised procedures, but almost invariably they will simplify to combinations of simpler procedures that belong to a small number of process types or 'families'. In this section we will be concentrating on idealised, mathematical descriptions of such 'basis' dopant infusion operations as they would appear in CMOS processes. This analysis is very generic by virtue of its parametric nature and helpful even in the absence of specific manufacturing process data that would allow ad hoc simulations by process-level tools of the TCAD (Technology Computer Aided Design) family (simulations that would be the next logical step after the purely theoretical calculations).

CMOS technology is based on creating doped area patterns on a p-type substrate. The reasons behind this choice of substrate have to do with the inherently larger mobility of free electrons in p-type silicon (a fact that speeds up n-MOS devices manufactures directly on top of substrate) and the fact that

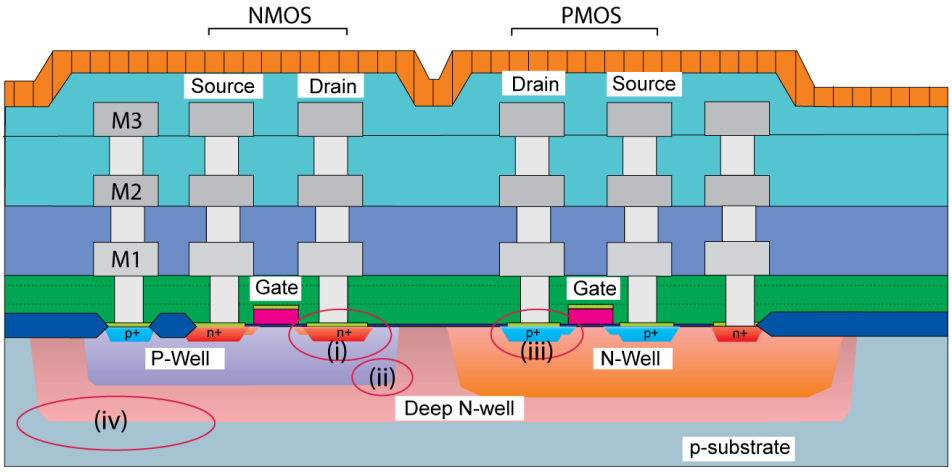

Figure 3.9: Cross section of a typical CMOS-fabricated die. Key elements are the available pn-junction types: (i): n+/p-well, (ii) p-well-n-well, (iii) p+/n-well, (iv) n-well/p-sub. The p-well/n-well junction is not alwys available. M1-M3 are different metallisation layers. Adapted from the XA035 technology datasheet available on the XFAB website. This is by no means an exhaustive list and includes structures such as p-wells that are not available in all CMOS technologies.

a p-type substrate can be connected to GND without risk of forward biasing any junction (for systems that operate in a GND-VDD power supply regime).

In terms of dopant infusion techniques, the CMOS manufacturing process can manipulate the front-end of a Silicon die by implanting ions deep inside the substrate by shooting them at the die with high kinetic energy, by diffusion of dopants into the substrate from an open window in the respective mask, or by diffusion of an ionic implant. The diffusion process will form 'diffusion' regions while ionic implantation will form 'well' regions, both of which can be of either n- or p-type polarity within a Silicon substrate. Diffusion of an ionic implant can create both diffusion and well regions. The basic mathematics of the processes behind the creation of diffusion and well regions will be examined in this section. A cross-section of a typical CMOS process is shown in Figure 3.9 for reference.

Diffusion and well regions can be combined to create the basic pn-junctions of CMOS: i) n-diffusion on substrate (n+/p-SUB). ii) p-diffusion on N-well (p+/NW). iii) N-well on substrate (NW/p-SUB). Here we assume a p-type substrate as is the norm with CMOS processes and in contrast to standard bipolar processes that are manufactured on an n-type substrate. Furthermore, the p-substrate can be either pre-doped by ion implantation (thus rendering it a very large, die size p-type well), or undoped, i.e in its 'native state'. This creates the further possibilities of: iv) n-diffusion on p-well (n+/PW) and v) N-well on P-well (NW/PW).

3.3.1 Dopant diffusion: planar process

A simple process of creating a diffusion region is the so-called 'planar process'. During this procedure a layer of oxide is first grown on the die and then patterned by use of photolithography to create so-called 'windows'. The oxide is itself called a 'mask'. Next, a layer of dopant material is spun on top of the entire die and the ensemble is heated at temperatures between 800-1200 ^oC [9]. As the ensemble stays hot, dopants start diffusing from the spun dopant coat into the Silicon substrate through the windows in the

mask. As time progresses the coat of dopants becomes thinner but remains of the same consistency. This yields the boundary condition that dopant concentration at the very surface of the die stays constant throughout the process.

Note: Of course as the diffusion process takes course, Silicon atoms also diffuse into the pure dopant since for every dopant atom entering the lattice a silicon atom is displaced. As a result, the Silicon-containing sub-volume of the material starts expanding upwards into the dopant. This layer may be removed when cleaning the die of dopant in subsequent steps but it being a superficial effect (affecting final doping concentrations primarily at the surface even though it does cause a departure from the boundary condition described below) we ignore it in this approximation for simplicity.

The geometric set-up for this section is a 1D trajectory that starts from the surface of the die ($x = 0$) and penetrates into the Silicon substrate perpendicular to the surface ($x > 0$). Diffusion processes follow Fick's 2nd law of diffusion:

$$\frac{dN_Z(x,t)}{dt} = D_Z \frac{d^2 N_Z(x,t)}{dx^2} \tag{3.91}$$

where $N_Z(x)$ is the concentration of dopant Z in Silicon and D_Z is the diffusion constant for dopant Z in Silicon. This equation describes the 1D case. The general n-dimensional case is given by:

$$\frac{dN_Z}{dt} = D_Z \nabla^2 N_Z \tag{3.92}$$

which is of the same form as the heat equation. The solution of this equation in 1 dimension under the boundary condition that throughout the entire duration of the diffusion process the concentration of dopants at the surface of the die ($x = 0$) is constant and that temperature is also constant (and therefore $D_Z = \text{const.}$) is:

$$\boxed{N_Z(x,t) = N_Z(0,t)\, erfc\left(\frac{x}{2\sqrt{D_Z t}}\right)} \tag{3.93}$$

where $N_Z(0,t) = const.$ as mentioned before and erfc(x) is the complementary error function. By inserting an appropriate value into the time variable a function describing doping concentration as a function of depth within the die ($N_Z(x)$) is formed. This can be used as $N_A(x)$ or $N_D(x)$ within the context of pn-junction physics as described in section 3.2.1.

Note: diffusion junctions consist of a main area, including all points well-away from the mask window edge and the 'edge regions' including all points that are sufficiently close to the edge of the mask window in order to be affected by 'edge' and 'vertex' effects. However, equation 3.93 only covers 1 dimension. When extended to 2D the implicit assumption is made that in the 2D-case the doping concentration along any pair of parallel lines crossing the die surface will be identical. As such, vertex and edge effects are not accurately modeled by equation 3.93. Nevertheless, the analysis will still hold approximately true for junctions occupying large surface areas. All of this is illustrated in Figure 3.10 while Figure 3.11

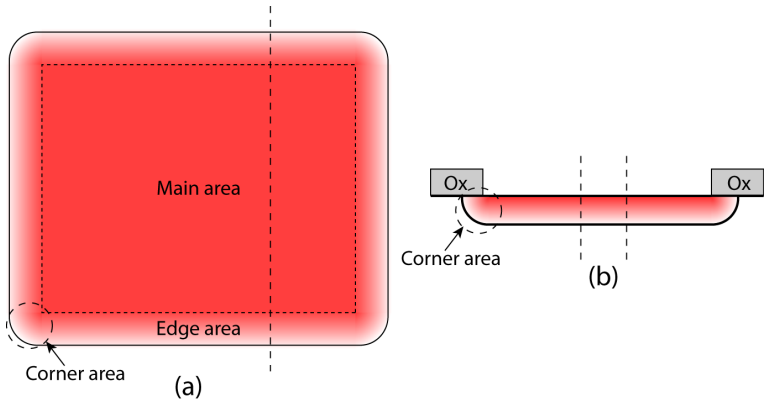

Figure 3.10: A diffusion region showing: (a) Top view with the main area, edge and corners annotated (where Fick's law does not apply without modifications). (b) Cross-section view taken across the dashed line in (a). 'Ox' denotes the oxide mask (not shown in (a)). The doping concentration along the parallel dashed lines in (b) is point-for-point exactly the same, as required for the generalisation of equation 3.93 from 1D to 2D. The intensity of the red colour in both panels loosely represents dopant concentration within the diffusion region.

shows qualitative plots of doping concentration for three different annealing times (diffusion time t from equation 3.93).

This highly simplified and idealised analysis ignores a number of non-ideal effects. According to [9] the diffusion process rarely occurs evenly throughout a 'window', diffusing dopants interact with the oxide/Silicon interface leading to doping concentrations near the surface no longer obeying the simple Fick law and sometimes diffusion regions may interact with each other or with other doped regions since the presence of 1 species of dopants may alter the diffusion rate of another. Thus, determining the precise doping profile of a diffusion region analytically is not a trivial task. According to [9] even powerful computers are finding it difficult to simulate diffusion processes numerically. Furthermore, the statistical variation of dopants in regions of very small size (of the order of a few millions of atoms within an orthogonal volume) introduces an additional layer of difficulties in estimating doping profiles for technologies with very small feature sizes.

3.3.2 Ion implantation

Ion implantation is a process where dopant atoms are pelted towards the Silicon die with known average kinetic energy and for a strictly specified interval of time. This process ideally creates a Gaussian doping profile in the direction perpendicular to the surface of the die ('vertical', or 'longitudinal' direction) and a rectangular window function in the direction parallel to the surface of the die (the 'lateral' direction). Patterned photoresist or poly-silicon masks (or a combination of both) can be used to control the spatial pattern of the implantation process. Ion implantation does not require high temperatures for processing, although a moderate temperature (800-900 ^oC) anneal at the end is performed in order to fix lattice damage caused during the procedure[5]. The annealing stage, however, does not fundamentally alter the properties of the doping profile function as evidenced by the fact that well-junctions with nearly vertical

[5]In many processes, however, the well is actively 'driven' deep into the substrate by deliberately subjecting the die to a high temperature anneal. Both the cases of a well created without this extra drive process and with it are discussed in the 'Art of analogue layout' [9].

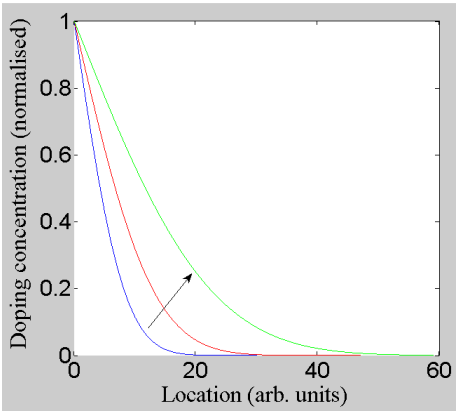

Figure 3.11: Normalised doping concentration vs. distance from the surface of the die (i.e. depth) obtained by implementing equation 3.93 in MATLAB. Concentration at $x = 0$ is kept at a constant value of 1 for the entire duration of the simulation. The blue trace indicates doping profile after 10 time unit steps, the red trace after 25 time steps and the green trace after 75 steps. The arrow shows the general trend over time.

side-walls can be created (see [9]) and that the so-called 'lateral straggle' effect, i.e. the effect that as dopant atoms penetrate the Silicon substrate and collide with lattice atoms they don't just slow down in the longitudinal direction, but are also diverted laterally, is predominantly inherent to the implantation procedure rather than the subsequent annealing (or at least with appropriate process parameter choice it can be so) [10].

The Gaussian profile in the longitudinal direction (perpendicular to the die surface) away from any well edges can be described by the following equation:

$$N_Z(x) = A_0 \exp\left(-\frac{(x - x_\mu)^2}{\sigma_x^2}\right) \tag{3.94}$$

where $N_Z(x)$ is the doping concentration of dopant Z as a function of depth only, A_0 is the peak concentration found at depth x_μ and depends on the implant dose, and σ_x describes the spread of the distribution. x_μ depends on the dopant energy. Here we have assumed that the ion implantation occurs perpendicular to the surface of the die, which means that the ions are shot towards the Silicon die at an angle of 0^o. This is most often not the case because the penetration depth of ion implants (parameter x_μ) does depend on the angle of implantation with respect to the Silicon crystal orientation. According to [9] a 7^o implant angle with respect to (100) Silicon is typical.

In reality creating a well featuring a single Gaussian doping profile allows too much doping concentration variability throughout the structure in the vertical direction, which is why typically ion implantation will consist of a series of individual implantation processes, each with its own dopant energy and duration parameters. The result of these processes can be approximated as a sum of Gaussians:

$$\boxed{N_Z(x) = \sum_i A_{0,i} \exp\left(-\frac{(x - x_{\mu,i})^2}{\sigma_x^2}\right)} \tag{3.95}$$

where i denotes the different implantation process routines employed and we assume that σ_x doesn't change after the application of each implantation procedure. $A_{0,i}$ denotes the peak implantation concentration for procedure i while $x_{\mu,i}$ similarly denotes the depth at which the maximum concentration occurs for procedure i, otherwise known as 'implantation depth'.

Although equation 3.95 shows the doping profile as a function of depth, it provides no clues as to how the concentration varies as one moves parallel to the surface of the die, i.e. it carries no information about lateral straggle. In one of the simplest models of straggle, the lateral distribution of dopants in a well structure can be expressed as a difference of complementary error functions [11, 12]:

$$N_Z(y) = \frac{B_0(x)}{2}\left[erfc\left(\frac{y-\zeta}{\sqrt{2}\sigma_y}\right) - erfc\left(\frac{y+\zeta}{\sqrt{2}\sigma_y}\right)\right] \tag{3.96}$$

where $B_0(x)$ is the 'steady state' doping concentration far away from the edge of the well at depth x, given by equation 3.94 (or equivalently the peak doping concentration at depth x), ζ is the 1/2-width of the implant mask opening, σ_y is a spread-like parameter that determines the inherent invasion extent of laterally displaced dopants into Silicon and the y coordinate will denote separation from the centre of the mask opening.

The overall 2D doping concentration $N_Z(x, y)$ is then given by replacing $B_0(x)$ with $N_Z(x)$ from 3.94 into 3.96:

$$N_Z(x,y) = \frac{A_0}{2}\exp\left(-\frac{(x-x_\mu)^2}{\sigma_x^2}\right)\left[erfc\left(\frac{y-\zeta}{\sqrt{2}\sigma_y}\right) - erfc\left(\frac{y+\zeta}{\sqrt{2}\sigma_y}\right)\right] \tag{3.97}$$

If we assume that lattice dislocations and dopants introduced by each implantation process don't significantly alter the effect of subsequent implantation procedures, then we can once again find an expression for multiple, successive implantations superposed on one another:

$$\boxed{N_Z(x,y) = \sum_i \frac{A_{0,i}}{2}\exp\left(-\frac{(x-x_{\mu,i})^2}{\sigma_x^2}\right)\left[erfc\left(\frac{y-\zeta}{\sqrt{2}\sigma_y}\right) - erfc\left(\frac{y+\zeta}{\sqrt{2}\sigma_y}\right)\right]} \tag{3.98}$$

A couple of solutions of equation 3.97 have been plotted in Figure 3.12 for different sets of the σ_x and σ_y parameters.

In this idealised model of an ion implantation process we have not considered a well-known effect that occurs at high implantation energies. As the implantation energy increases, the Gaussian distribution in the vertical (longitudinal) direction tends to become skewed, presenting a long tail towards the surface and steep roll-off towards the deep end of the die. Such modeling issues can be solved if one is to use Pearson distributions [13, 14], although the works cited tend to concentrate on shallow implants ($< 1\mu m$) where the effects of the Gaussian skew are less pronounced.

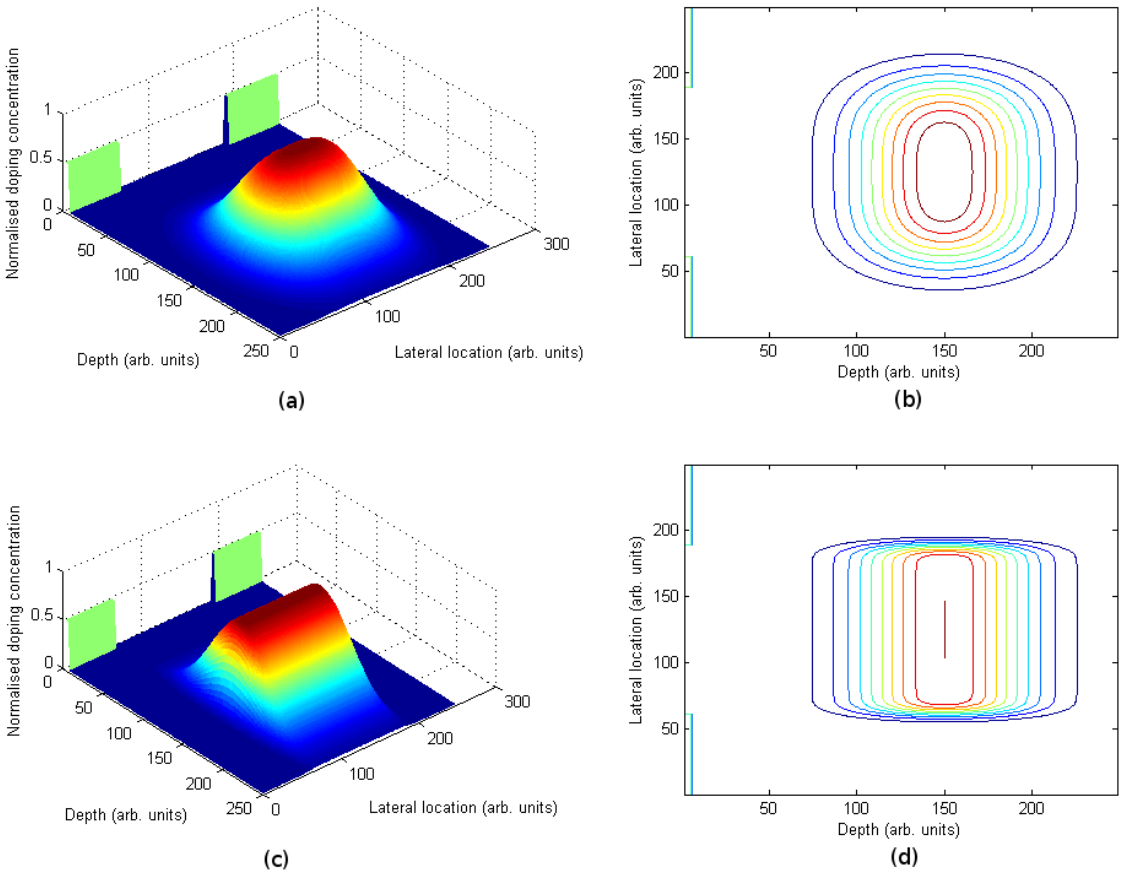

Figure 3.12: Simulated ion implantation process shown via MATLAB-generated surface and contour plots for normalised doping concentration vs. location. Location coordinates are in arbitrary units of the same scale. The left half-plane of the image (a,c) shows surface plots, the right half-plane (b,d) shows contour plots, the top half-plane (a,b) shows results for a process with $\sigma_x = 50$ arbitrary units and $\sigma_y = 20$ arbitrary units and the bottom half-plane (c,d) the same results for a process with $\sigma_x = 50$ arbitrary units and $\sigma_y = 5$ arbitrary units. Both processes have an implantation depth (x_μ) of 150 arb. units. The green rectangles at $x \approx 0$ visible in the surface plots denote the spatial extent of the mask used for the simulation of both processes. The mask is also visible in the contour plot (green lines close to $x = 0$).

Combination of implantation and diffusion processes

A crucial effect that befits study concerns the combination of an implantation step with a diffusion step, in that order. This effect is crucially important because diffusion regions in CMOS can be created by a combination of a shallow implant + diffusion whilst well regions tend to be created by a deep implantation and a long and slow anneal [9]. Thus in any realistic CMOS technology both diffusion and well regions will exhibit characteristics that do not quite match their ideal, individual manufacturing processes.

The 1st step in creating the structure that we shall hence-forth call a 'diffused well' is a standard ion implantation process. The analysis for this process leads to equation 3.97 or, for a cascade of wells equation 3.98. The doping function resulting from the initial implantation procedure will be denoted as $I(x, y)$ Next, this needs to be 'diffused'. For this we need to revisit Fick's law (equation 3.92) in its n-dimensional vector form, since we are diffusing a function of 2 variables, x and y. The solution to Fick's law now has a fundamentally different boundary condition:

$$N_Z(x, y, 0) = I(x, y) \tag{3.99}$$

where N_Z is now a function of x, y and time t. Given this initial condition, the solution to Fick's law takes the form (generalisation from [1]):

$$\boxed{N_Z(\vec{x}, t) = \int_{\mathbb{R}^2} I_\delta(\vec{x} - \vec{\tau}, t) I(\vec{\tau}) d\vec{\tau}} \tag{3.100}$$

where $\vec{x} \equiv (x, y)$ and $\vec{\tau}$ is a dummy variable used for the integration. Crucially, $I_\delta(\vec{x}, t)$ is the solution of the 2D equation with the initial condition being a Dirac delta function $\delta(\vec{x})$. The solution for such initial condition is given by:

$$I_\delta(\vec{x}, t) = \frac{1}{4\pi D_Z t} \exp\left(-\frac{\vec{x}^2}{4 D_Z t}\right) \tag{3.101}$$

where $\vec{x}^2$ denotes a dot product.

Equation 3.100 has been numerically solved with MATLAB for a shallow and a deep implantation process (Figure 3.13). Notably, both implantations experience significant pattern blur after the diffusion process, but in the case of the shallow implantation the surface of the die ends up with a high concentration of dopant atoms in way reminiscent of the planar diffusion process and the region of maximum normalised doping concentration shifts towards the deep end of the material because of losses at the surface. We have assumed that any dopants reaching the surface are freed from the Silicon crystal and therefore $N_Z(x, y, t) = 0$ for any values of y and t if $x < 0$ i.e. the boundary condition is that outside the Silicon crystal the doping concentration is zero.

Equation 3.100 also has severe limitations: once dopant atoms diffuse to the surface of the die, the diffusion process meets new boundary conditions. Dopant atoms cannot carry on diffusing into the

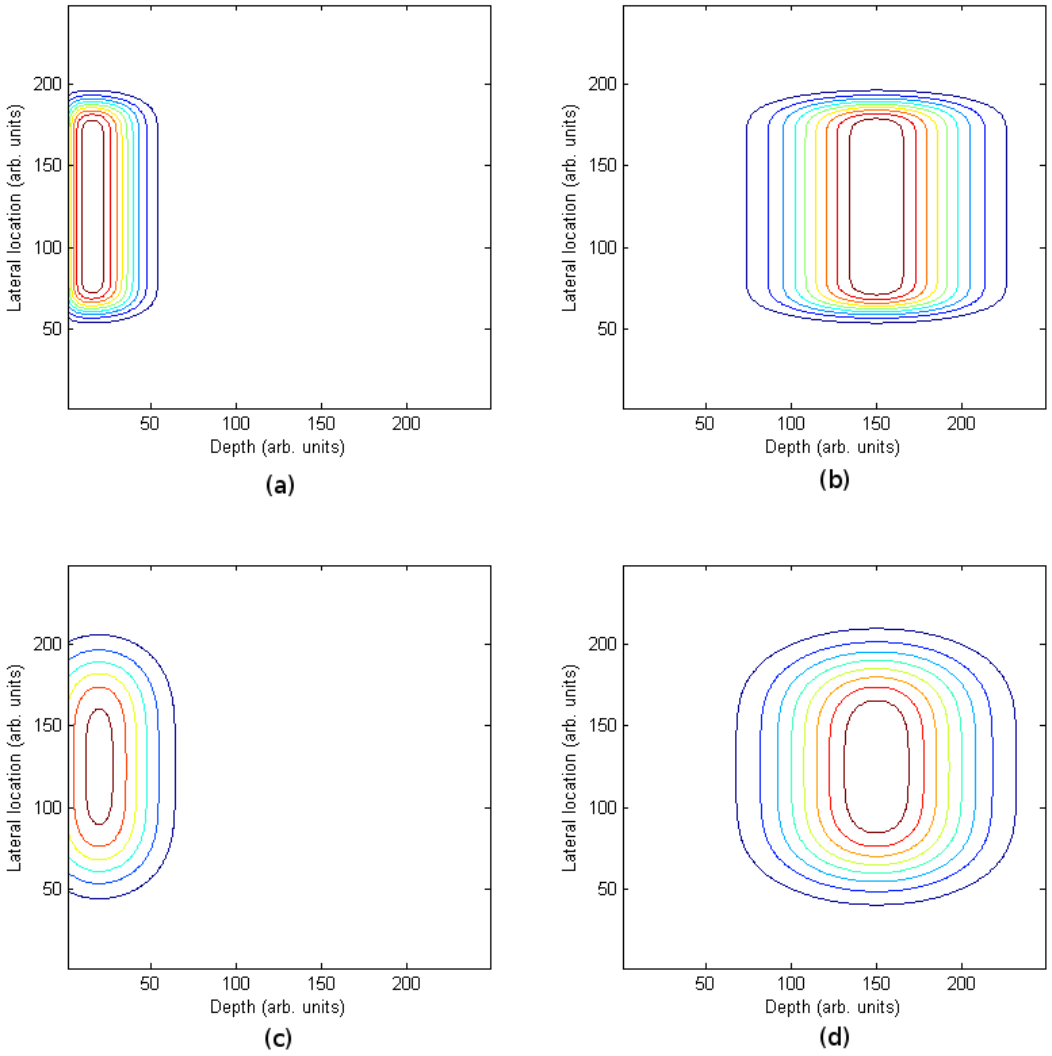

Figure 3.13: Simulated ion implantation processes followed by diffusion (high temperature anneal) shown via MATLAB-generated contour plots for normalised doping concentration vs. location. Colours towards the red end of the visible spectrum indicate high values, while colours towards the violet end indicate low values. Location coordinates are in arbitrary units of the same scale. The left column (a,c) shows the case where the implantation is 'shallow', the right column (b,d) the case where the implantation is 'deep'. The top row (a,b) shows dopant distribution after the implantation process and the bottom row (c,d) shows dopant pattern after 15 time steps during the subsequent diffusion process. Units of distance in both x- and y-directions are arbitrary, but of the same scale. Doping concentrations are normalised and therefore the areas of maximum normalised doping concentration in the bottom row feature substantially lower absolute values for dopants per unit volume. The parameters used for this simulation were: Shallow implantation: Longitudinal straggle spread = 25 arb. units. Implantation depth = 15 arb. units. Deep implantation: Longitudinal straggle spread= 50 arb. units. Implantation depth = 150 arb. units. Common: Lateral straggle spread coefficient = 5 arb. units. Diffusion constant = 10 square arb. units per unit time.

Silicon oxide covering the die front-end as if it were crystalline Silicon, which fundamentally changes the distribution near the surface compared to what free diffusion in an infinite medium would indicate.

3.3.3 Diffusion and implantation summary and comments

In summary we have shown simplified, ideal models for diffusion and well regions that relate doping concentration to location on a cross-sectional sheet of semiconductor. The generalisation to 3 dimensions can be made trivially if edge and vertex effects can be assumed to be negligible.

For diffusion regions constructed via the planar diffusion process we have found an analytical expression for doping concentration vs. depth and ignored vertex and edge effects (the perimetric fringe of the diffusion area).

For well regions we have shown a simple function that describes doping concentration as a function of both depth and lateral displacement from the centre of the mask gap. Here we have ignored regions where the basal facet of the well (parallel to the surface of the die) meets 'side-wall' facets (perpendicular to the surface of the die), as well as regions where side-wall facets meet each other, but we have not ignored the side-wall facets themselves, thus allowing us to differentiate between the basal junction and side-wall junctions formed between the well and the substrate. The 'Gaussian skew' problem for high energy implantations has also been ignored here.

Finally, we have considered diffused wells where an ion implantation is followed by a diffusion process and found an equation that can be used to predict doping concentration as a function of location and the mathematical parameters of the applied manufacturing steps. We have ignored surface effects in this analysis.

Both diffusion and implantation procedures are characterised by manufacturing parameters, which then define the mathematical parameters that fit into their model equations. The rough equivalence between mathematical modeling and manufacturing parameters for diffusion and implantation can be seen in Table 3.1.

3.4 Phototransduction

Phototransduction can be achieved, amongst other ways, by use of a photodiode, which is a pn-junction operating in the power generating 4th quadrant of the I-V plane[6]. The final objective is to convert as many incident photons into an electric current. Metal-semiconductor (Schottky-based) photodiodes also exist, but their study is outside the scope of this thesis. The analysis in this section it split into: i) Basic electro-optical characteristics (e.g. annotated I-V curves). ii) Parasitic effects and manufacturing factors.

[6]On an I-V plane we have four quadrants, each referring to a unique combination of current and voltage signs. In the case of two-terminal devices, each point in the I-V plane determines an operating point where V is the voltage across the device and I the current flowing through it. Operating points lying in quadrants I and III are 'dissipating' since a positive voltage leads to a positive current and a negative voltage to a negative current (current flows towards the lower voltage node). Points in quadrants II and IV, on the other hand, are 'generative' as positive voltages cause negative currents and vice versa (current flows towards the higher voltage node).

Table 3.1: Correspondence between mathematical and manufacturing parameters for doped regions created by planar diffusion and ionic implantation. Quick variable guide: i) D_x is read as: diffusion constant of dopant x in Si. ii) The temperature schedule reflects finite cooling and heating times. iii) Surface concentration can also be a function of location on the plane of the die surface, but is termed $N(0)$ to reflect the assumption that this is not the case. The '0' refers to the depth variable. iv) ε_x is read as: average energy of each dopant ion of type x, where x denotes an element (P, Sn, B etc.). v) σ_x is read as: standard deviation in the longitudinal direction.

DIFFUSION (PLANAR)

Mathematical variables		**Manufacturing variables**	
Diffusion constant:	D_x	Temperature schedule:	$T(t)$
Surface concentration:	$N(0)$	Surface concentration:	$N(0)$
Diffusion time:	t_{diff}	Diffusion time:	t_{diff}

IMPLANTATION (LONGITUDINAL)

Mathematical variables		**Manufacturing variables**	
Mean depth:	x_μ	Ion energy:	ε_x
Standard deviation:	σ_x	Implantation angle:	θ_i
Peak concentration:	N_{peak}	Implantation dose:	M_{dose}

3.4.1 Basic electro-optical characteristics

The principle of operation of a photodiode revolves around the phenomenon of interband absorption described in section 3.1.1. Specifically, when a photon causes an electron-hole pair generation within the depletion region of the photodiode the electric field between bulk-n and bulk-p serves to separate them and drive the free electron into the n-type region and the hole into the p-type thus creating a new type of current called 'photocurrent' and contributing energy to the system. A number of points need to be made:

1. Photons of energy lower than that of the band-gap E_g will not interact with the device except in the unlikely case where they cause an extrinsic transition or in cases where two photons work in synergy to create a transition (akin to what happens in 2- and multi-photon microscopy). Photons with energy above E_g will waste any excess energy $E_{ph} - E_g$, where E_{ph} is the energy of an incident photon, in the form of heat.

2. The separated electron-hole pair increases the potential across the pn-junction above V_{bi}, so after a certain amount of time the diffusion/drift current balance within the pn-junction will restore equilibrium.

3. Electron-hole pairs are routinely generated throughout the entire volume of any semiconductor material. However, in the absence of the relatively strong electric field, such as that present in the depletion regions of semiconductor junctions the pair will tend to quickly recombine. For that reason we only consider electron-hole pair generation in depleted material as 'useful'.

In the case where a certain amount of electron-hole pairs are generated by incident photons every second we may model this as an additional current flowing throughout the junction alongside the diffusion

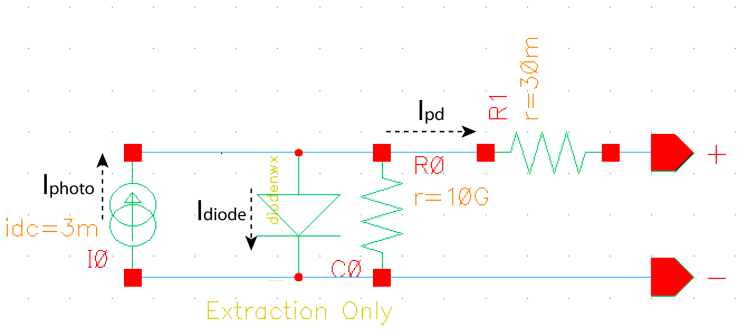

Figure 3.14: Diode macromodel showing the pn-junction (C0) with its associated shunt resistance (R0) and series resistance (R1), as well as a current source representing the photocurrent (I0) under constant illumination. The values indicated next to the respective components are typical for integrated solar cells.

and drift currents that have been studied in section 3.2.1. At this point we assume that the existence of a non-zero photocurrent does not affect the manifestation of the drift and diffusion currents significantly. This assumption will generally hold since photocurrent is based on the electron-hole pair generation process that creates free carrier pairs out of valence electrons and adds them to the population of free electrons stemming from the conduction band of n-type semiconductor material. A similar phenomenon is observed for free holes. It is these 'original' free carriers partake to drift and diffusion currents normally.

The photocurrent can then be modeled as a current source connected in parallel with a regular diode. A macro-model of a photodiode can be seen in figure 3.14, where the shunt and series resistances of the pn-junction have also been included. The shunt resistance is typically of very high value and is only significant at those reverse bias voltages where the impedance of the diode component of the device is comparable to it[7]. The series resistance reflects mostly the additional resistance seen at the terminals of a real device by external components and is not examined here. The anode and cathode terminals are where any external device would connect to the overall set-up.

Importantly, we notice that in the simplified macromodel where the series and shunt resistances are ignored the photocurrent has no option, but to flow through the diode element. In response, the diode element becomes forward biased and assumes the correct bias voltage that sinks all of the produced photocurrent. Connecting a load across the structure will then introduce an impedance which will compete with the diode for photocurrent, lowering terminal voltage across the device but powering the load in exchange (current-divider action).

The ability of a photodiode to provide electrical energy can be seen in its I-V characteristic, which can be described by:

$$\boxed{I_{pd} = I_{diode} - I_{photo} = I_s \left(\exp\left(\frac{qV_{pd}}{kT} \right) - 1 \right) - I_{photo} = I_s \left(\exp\left(\frac{V_{pd}}{V_T} \right) - 1 \right) - I_{photo}} \tag{3.102}$$

where I_{pd} is the current flowing across the photodiode structure, I_{diode} is the current flowing across the current flowing across the ideal diode component of the photodiode structure, V_{pd} the corresponding

[7]The shunt resistance will typically be some extreme parasitic current path between the terminals of the device typically arising from physical defects of the device, such as Silicon dislocations, unfavourably placed grain boundaries in malformed Silicon etc. Consequently, it is most often ignored in photodiode models. We mention it here for completeness.

voltage across the device, I_s is the reverse saturation current and depends on its area, $V_T = \frac{kT}{q}$ is the thermal voltage and I_{photo} is the photocurrent.

The reverse saturation current is given in Sze [1] as:

$$I_s = qA\left(\sqrt{\frac{D_h}{\tau_h}}\frac{n_i^2}{N_D} + \sqrt{\frac{D_e}{\tau_e}}\frac{n_i^2}{N_A}\right) \tag{3.103}$$

where A is the area of the junction and τ_e and τ_h are the electron and hole carrier lifetimes respectively. The carrier lifetimes are the coefficients that describe the exponential decay of free carriers from a non-equilibrium concentration to an equilibrium concentration. The details of how these parameters are determined can be found in [15].

The following observations are made: a) The I-V function is monotonically increasing. b) At $V_{pd} = 0$ we get $I_{sh} = -I_{photo}$, where I_{sh} is the short circuit current. c) In order to obtain $I_{pd} = 0, V_{pd}$ must satisfy:

$$\boxed{V_{oc} = V_T ln\left(\frac{I_{photo}}{I_s} + 1\right)} \tag{3.104}$$

where V_{oc} is the open circuit voltage.

In general the photocurrent will be significantly larger than the reverse saturation current and therefore $V_{oc} \gg 0$. This proves mathematically that the device operates in the 4th quadrant. An example of an I-V plot of a photodiode is provided in Figure 3.15.

If we trace a route from I_{sh} to V_{oc} along the I-V characteristic, we will obtain an infinite set of points with coordinates $(I(V), V)$ at which the system contributes power i.e. the following holds:

$$P(I, V) = IV < 0 \tag{3.105}$$

where $P(I, V)$ represents power dissipation through a 2-terminal black box as a function of current and voltage drop across the terminals of the box (and hence $P < 0$ implies power generation).

At least one of the $(I(V), V)$ points will represent a global minimum for the function $P(I, V)$. That point will represent the maximum power operating point of the photodiode. We can find the pair corresponding to the maximum power point numerically, but first we need to find a suitable equation to submit to the numerical solver. If we ignore shunt and series resistances, we begin by setting $I = I_{pd}$, $V = V_{pd}$ in 3.105 and then substituting I_{pd} from equation 3.102. We obtain:

$$\boxed{P(I, V) = V_{pd}I_s \exp\left(\frac{V_{pd}}{V_T} - 1\right) - I_{photo}V_{pd} = I_sV_{pd}\exp\left(\frac{V_{pd}}{V_T}\right) - (I_s + I_{photo})V_{pd}} \tag{3.106}$$

Differentiating to obtain dP/dV we get:

$$\frac{dP}{dV_{pd}} = I_s \exp\left(\frac{V_{pd}}{V_T}\right) + \frac{I_sV_{pd}}{V_T}\exp\left(\frac{V_{pd}}{V_T}\right) - (I_s + I_{photo}) \tag{3.107}$$

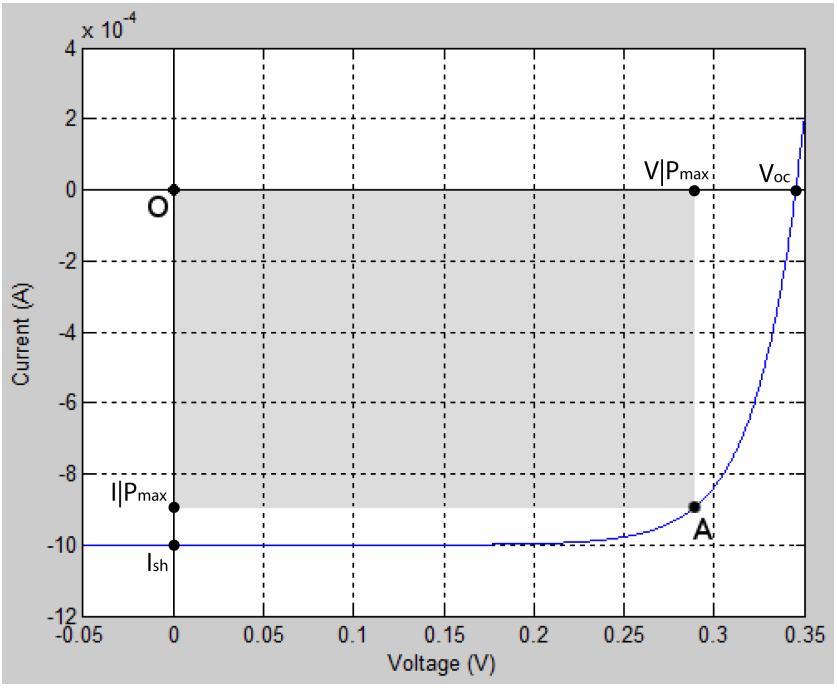

Figure 3.15: I-V characteristic curve of an ideal photodiode. The reverse saturation current has been set to the value of 1nA and the photocurrent has been assumed to be 1mA. The area of the gray rectangle corresponds to the power provided by the photodiode at the maximum power operating point as defined by point A. Also shown are: The short circuit current (I_{sh}), the open circuit voltage (V_{oc}) and the current and voltage at the maximum power point ($I|P_{max}$ and $V|P_{max}$).

Since dP/dV is a monotonically increasing, continuous and differentiable function of V_{pd}, it can only have 1 zero-crossing and therefore the global minimum is a unique point that appears at the value of V_{pd} where $dP/dV = 0$. Equating the right hand side of 3.107 to zero yields the following implicit relation:

$$V_m = V_T ln\left(\frac{I_s + I_{photo}}{I_s}\right) + V_T ln\left(\frac{V_T}{V_m + V_T}\right) = V_{oc} + V_T ln\left(\frac{V_T}{V_m + V_T}\right) \tag{3.108}$$

where V_m is the voltage at the maximum power point.

Once the value of V_m is calculated it can be substituted in 3.102 to yield the photodiode current at the maximum power point (I_m) and hence the maximum power output can be computed.

Another key metric of photodiodes is the power conversion efficiency η, defined as the ratio of electrical power output to optical power input. At the maximum power point:

$$\eta = \frac{I_m V_m}{P_{opt}} \tag{3.109}$$

where P_{opt} is the optical power input.

The fill-factor F is defined as the ratio of the maximum power output of the photodiode over the product $I_{sh}V_{oc}$:

$$F = \frac{I_m V_m}{I_{sh} V_{oc}} \tag{3.110}$$

Using 3.110 in 3.109 yields an alternative form of 3.109:

$$\boxed{\eta = \frac{F I_{sh} V_{oc}}{P_{in}}} \tag{3.111}$$

Finally, photodiodes can be assessed by use of, amongst others, the concepts of external and internal quantum efficiencies (EQE and IQE respectively). The internal quantum efficiency is defined as the ratio of the number of photons absorbed by the photodiode to the number of electron carriers created whilst the external quantum efficiency is the ratio of the number of photons incident upon the surface of the photodiode (regardless of whether they are absorbed or not) to the number of generated electron carriers. The units of both metrics are electrons/photon.

IQE is affected more by the physics of photodiodes and can be related to the discussion about inter-band absorption done in section 3.1.1, whilst EQE will also be affected by the manufacturing and packaging of the device as well.

3.4.2 Parasitic effects and manufacturing factors

Parasitic series resistance: The effects of a load series resistor can be expressed with the help of equation 3.102:

$$I_{photo} = I_{pd} + I_{load} = I_s \left(\exp\left(\frac{V_{pd}}{V_T} \right) - 1 \right) + I_{load} \tag{3.112}$$

where I_{load} is the current sunk by the load, V_{pd} is the voltage across the ideal component of the photodiode.

A MATLAB plot pertaining to the photodiode characterised in Figure 3.15 (same I_s) can be seen in Figure 3.16. Power delivered to the load has been plotted against load impedance and photocurrent. The simulation is numeric and has resulted into a characteristic 'shark-fin' surface. Notably, parasitic series impedances of a few tens of Ohms in magnitude will not necessarily influence the power delivery to the load dramatically if the load itself lies near the optimum region. The reason becomes clear if we consider the parasitic resistance as a small, unsolicited increase in nominal load impedance. Specifically, Figure 3.15 shows that changes below $10\,\Omega$ do not seem to affect power transmission efficiency dramatically at the peak of performance (as would be the case in a system with an MPPT (Maximum Power Point Tracking) module. At load impedances below $200\,\Omega$ greater gradients are observed (approx. $100\,\mu W/100\,\Omega$; corresponding plot not shown.), which with respect to the maximum achievable power output from the photodiode given a photocurrent magnitude still means that parasitic resistances of a few Ohms will not have a dramatic effect on power transmission. How well this generalises to simulated systems with different photodiode characteristics is a matter of further study.

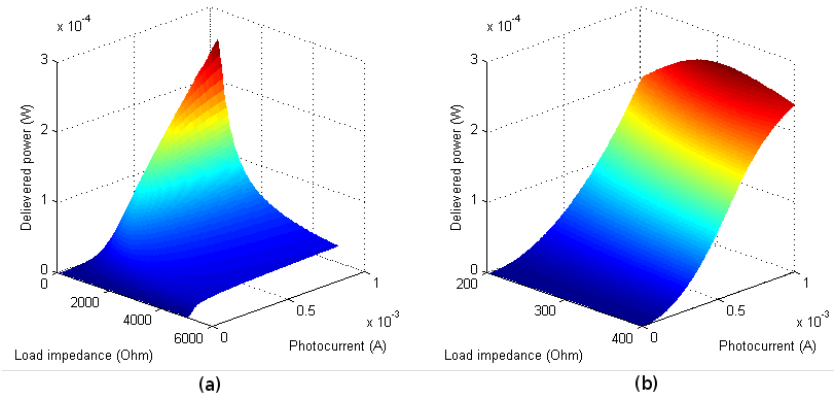

Figure 3.16: Power delivery to a load from a photodiode with a reverse saturation current of 1 nA has been numerically simulated as a function of load resistance and photocurrent. Shown are: (a) the full simulation space, and (b) a zoom on the crest of the 'shark fin' shape of panel (a).

Recombination current: Generation of electron-hole pairs is constantly counteracted by recombination. Both phenomena occur in both depleted and un-depleted semiconductor material, but the reason the depletion region contributes most of the photocurrent is that the strong electric field in depleted Silicon separates the pair before it has a chance to recombine. Thus, in bulk-Si, where the electric field is weak, intense recombination destroys almost all generated electron-hole pairs while in the depleted region recombination is kept at much lower levels; sufficiently low to allow significant net electron-hole pair formation.

For the core, photocurrent-generating depleted region of the pn-junction a formula that describes the recombination current I_{rc} occurring via single-level recombination centres is given by Sze [1]:

$$I_{rc} = I_s^* \left(\exp\left(\frac{V}{2V_T} \right) - 1 \right) \tag{3.113}$$

where the reverse saturation current I_s^* now takes the form:

$$I_s^* = qA \frac{n_i W}{\sqrt{\tau_e \tau_h}} \tag{3.114}$$

Manufacturing fill factor: Finally, an important manufacturing consideration regarding photodiodes is quantified by a parameter called also 'fill factor', but this time representing the ratio of the active area of a photodiode, typically a pixel in a digital camera, over the entire area of the photodiode. To differentiate from the power-efficiency related fill factor term used above we shall dub it the 'manufacturing fill factor' (F_M).

In the case of a simple photodiode, the F_M can tell us how much of the pn-junction is shaded by metallisation and contacts. This can be more significant than it appears as the density of metallisation will determine to a certain extent the effective parasitic R_s of the device for instance. Free carriers need to traverse bulk (doped) silicon until they reach the ohmic contacts to the wires that connect them to their respective loads. The length of their trajectories can affect the resistance they 'see' along the way. This parasitic R_s in combination with the junction capacitance C_j will play a determining role when it

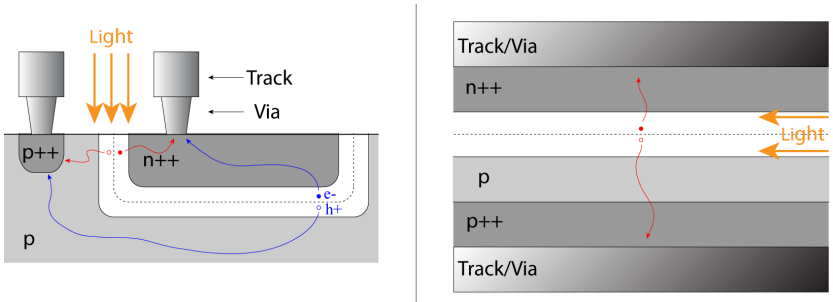

Figure 3.17: Comparison between two physical configurations of photodetector pn-junctions with respect to typical electron and hole trajectories from some random eletron-hole pair generation site to the nearest back-end contacts. (a): 'traditional' phototransducer pn-junction and (b): 'fast' phototransducer. In the traditional case a couple of electron-hole pairs are shown. The red pair 'sees' low resistance between its generation location and the metal contacts while the blue pair 'sees' high resistance. Large phototransducers with scant metallisation will give rise to many 'high-resistance' electron-hole pairs and thus will behave as slow (high inherent RC) devices. In the fast case, all generated electron-hole pairs are of the low resistance variety. White areas between p- and n-type semiconductor areas represent depleted Silicon and the dashed line represents the metallurgical surface.

comes to the specification of the photodiode's high frequency cut-off point. Thin junctions with 100% metallisation coverage either side will be fast, but useless for phototransduction unless the light falls upon them from a direction belonging to the metallurgical surface plane, for example (see Figure 3.17).

3.5 Conclusions

Throughout this chapter we have laid the basic theoretical principles that the entire thesis is based on. We have seen that electro-magnetic radiation can be absorbed by semiconductors by use of a couple of phenomena, namely inter-band absorption and free-carrier absorption, each operating in different wavelength ranges. Inter-band absorption is useful for processing high energy visible light and turning it into power whilst free-carrier absorption is a weaker phenomenon that only becomes significant in the low energy IR part of the spectrum, where any inter-band absorption phenomena have waned.

In terms of both FCA and power scavenging we have seen the pivotal importance of doping concentration in the vicinity around the metallurgical surface to performance. For purposes of modulation it was shown that high doping concentrations and abrupt changes between p- and n-type regions improve performance. The mathematical description of the issue reveals that when modulating the width of the depletion region two counteracting effects come into play: first high doping concentration at the modulated volume tends to linearly increase the effect of modulation on light passing through the junction (modulation depth is proportional to doping concentration within the modulated volume) whilst secondly, the inherent shrinkage of the modulated volume is proportional to the square root of the doping concentration. The net effect is that higher doping concentration increases modulation depth at a square root rate. For the purposes of power scavenging, on the other hand, low doping and wide depletion regions are required. This is because only electron-hole pairs generated within depleted Silicon exhibit a high probability of being separated and swept away by the depletion zone electric field before suffering recombination; a well-known and studied phenomenon.

The crucial role of appropriate biasing in determining modulator efficiency was also examined in this chapter. It was determined that reverse bias needs to be used in order to avoid flooding the depletion region with transiting carriers and that the bias voltage difference between unmodulated and modulated states (on and off states) yields a sub-square root improvement in modulation depth.

Moreover, having identified the importance of doping concentration profile for the function of pn-junctions we then examined some theoretical and yet reasonably realistic doping profiles as would be created by idealised CMOS manufacturing processes. Diffusion and implantation regions have been tackled as they form the basis of CMOS front-end manufacturing. Important learnings derived from this exercise imply that diffusion and well to substrate side-wall junctions are likely to be prime candidates for efficient optical modulation because of the high doping concentrations involved and crucially the abruptness of the transition between p- and n-type regions. Furthermore, an examination of the parameters that influence doping profile reveals that smaller feature size technological nodes are likely to perform better in the capability of modulators. This occurs because shrinking the technology tends to increase both doping concentrations and the abruptness of the transitions between p- and n-type material. These observations hint towards what might be efficient modulator design and were taken into account when designing the modulator test structures for this work (See appendix A for the full design repository).

Thus we have laid all the groundwork upon which we can begin to build a theoretical model of our proposed electro-optical communications platform. This is the subject of a dedicated main thesis chapter (no. 4). The models developed therein rely heavily on theory from this chapter, modified and adapted to help characterise realistic pn-junctions featuring doping profiles and concentrations that bear some semblance to those manufactured in standard CMOS technology.

Bibliography

[1] S. M. Sze, *Physics of Semiconductor Devices.* Wiley, 2nd ed., 1981.

[2] A. Serb, K. Nikolic, and T. G. Constandinou, "A CMOS-based light modulator for contactless data transfer: theory and concept," *SPIE proceedings*, vol. 7943, pp. 794317–794317–12, 2011.

[3] e. a. R. P. Feynman, *The Feynman lectures on physics.*, vol. 2. Addison-Wesley, 1964.

[4] M. A. Green, "Self-consistent optical parameters of intrinsic silicon at $300k^o$ including temperature coefficients," *Solar Energy Materials and Solar Cells*, vol. 92, no. 11, pp. 1305 – 1310, 2008.

[5] D. Schroder, N. Thomas, and J. Swartz, "Free carrier absorption in silicon.," *IEEE journal on solid state circuits*, vol. SC13, pp. 180–187, 1978.

[6] R. A. Soref and B. R. Bennett, "Electrooptical effects in silicon," *IEEE journal of quantum electronics*, vol. QE-23, no. 1, pp. 123–129, 1987.

[7] L. Keldysh, "Behavior of non-metallic crystals in strong electric fields," *Soviet Journal of Experimental and Theoretical Physics*, vol. 6, p. 763, 1958.

[8] T. Constandinou, *Photodiodes in modern deep sub-micron* CMOS *technology.* PhD thesis, Imperial College, South kensington Campus. Chapter 5.

[9] A. Hastings, *The art of analog layout.* Pearson Prentice Hall, 2006.

[10] R. Williams and M. Cornell, "The emergence impact of dram-fab reuse in analog and power-management integrated circuits," in *Bipolar/Bi*CMOS *Circuits and Technology Meeting, 2002. Proceedings of the 2002*, pp. 45 – 52, 2002.

[11] R. Oven and D. Ashworth, "A comparison of two-dimensional ion-implantation profiles," *Journal of Physics D: Applied Physics*, vol. 20, no. 5, p. 642, 2000.

[12] R. Oven, D. Ashworth, and M. Bowyer, "Formulae for the distribution of ions under an ideal mask," *Journal of Physics D: Applied Physics*, vol. 25, no. 8, p. 1235, 2000.

[13] A. Tasch, H. Shin, C. Park, J. Alvis, and S. Novak, "An improved approach to accurately model shallow B and BF_2 implants in silicon," *Journal of The Electrochemical Society*, vol. 136, no. 3, pp. 810–814, 1989.

[14] K. Suzuki and R. Sudo, "Analytical expression for ion-implanted impurity concentration profiles," *Solid-State Electronics*, vol. 44, no. 12, pp. 2253–2257, 2000.

[15] J. Pankove, *Optical processes in semiconductors.* Dover Publications, 2010.

Chapter 4

Theoretical treatise of proposed electro-optical communications system

Before attempting to construct the proposed system, a certain amount of theoretical modeling had to be carried out in order to determine how the background physics (discussed in chapter 3) would practically work within the context of CMOS electronics. The same chapter also includes background information on idealised doping processes in CMOS technologies. However, important theoretical developments directly affecting our work in electro-optical modulation would include finding out the characteristics of realistic pn-junctions and computing the optical pathway from emitter to detector under various possible, realistic opto-mechanical configurations of the system. This chapter attempts to address these objectives by means of what we shall call 'beam fascicle analysis' of an electro-optical modulator set-up.

'Beam fascicle analysis' is essentially a form of finite element analysis suited for broad fan-out electromagnetic radiation. It works by breaking the beam apart into small fascicles that depart the emitter within a very tight range of angles and then analysing their behaviour throughout the entire optical path from emitter to detector. This analysis is performed for three possible modulator configurations: transmittance mode, reflectance mode and bounce mode. Then theoretical results are shown that could be fed into a numerical solver in order to provide an exact theoretical solution to the idealised configuration problems. The configurations are thoroughly explained in the main body of this chapter.

This chapter is organised as follows: First, the practicalities of electro-optical modulation in a CMOS-like setting are examined (4.1). Then, a couple of worked examples will be given where electro-optical modulation performance has been theoretically calculated for an ideal abrupt junction case (4.2) and for a more realistic CMOS pn-junction (4.3). Finally, a short summary will place the learnings of the chapter into perspective (4.4).

4.1 Electro-optical modulation: Implementation in CMOS

In this section we shall deal with the engineering aspect of free-carrier absorption-based modulation. On the modulation side of the system, the aim is to modulate the intensity of an EM beam as 'deeply' as

possible i.e. to maximise the ratio $\frac{|I_{mod}-I_{unmod}|}{I_{unmod}}$ (a parameter we shall call 'fractional modulation depth'), where I_{mod} is the intensity of the beam when the modulator is on and I_{unmod} when the modulator is inactive. Similarly, we wish to maximise the value of $|I_{mod} - I_{unmod}|$ (the 'absolute modulation depth). These key performance metrics will be determined by the geometry of each device, the manufacturing process used to create it and the choice of operational parameters.

4.1.1 Free carrier-based optical amplitude modulation

In this section, we consider key engineering decisions associated with the implementation of electro-optical modulation in standard CMOS. These concern primarily the wavelength of operation and modulator geometry.

In terms of operational wavelength, we need to revisit Shroeder and Soref's equations on the intensity of free-carrier absorption:

$$\alpha = \left(\frac{e^3\lambda^2}{4\pi^2c^3\epsilon_0 n}\right)\left[\frac{n_e}{m_e^2\mu_e}\right] \tag{4.1}$$

where λ is the incoming radiation wavelength, n_e is the concentration of free electrons, m_e^2 the electron effective mass, μ_e the electron mobility, n the refractive index and the rest of the parameters are the usual universal constants.

It becomes clear that the most easily engineerable parameter is the wavelength and that higher wavelengths will dramatically improve modulator performance. Choosing a suitable wavelength will be determined by physical and practical considerations; ideally we would like free-carrier phenomena to dominate. As an absolute lower bound we have the approx. 1.1 micron wavelength below which Silicon becomes opaque due to inter-band absorption. In practice a lower bound of 1.3 microns is safely above the limit where inter-band effects become insignificant. As an upper bound, there is no obvious physical reason posing a boundary. However, practical considerations such as the availability of good quality photodetectors, possible refractive phenomena created by the metallisation of the die back-end and the availability of good quality emitters will impose 'soft' limits. We recommend the 1.8 micron wavelength as a suitable upper bound as reasonably cheap Germanium photodetectors can easily operate within the full 1.1-1.8 micron range. As a result of this assumption we shall be completely ignoring inter-band absorption in the analysis performed in this section. It must be also noted that broad-band illumination is equally acceptable in principle, the specifics being determined by the application at hand.

In terms of geometry, three configurations of a proposed optical modulator set-up will be described conceptually. Hence forth they will be referred to as the 'transmittance' modulator, the 'reflectance' modulator and the 'bounce' modulator structures. Each of these configurations comes with increasing technical complexity but also enhanced modulation capability and all of them are illustrated in Figure 4.1. The principle of operation in all configurations revolves around generating an EM beam off-chip, modulating it on-chip and detecting it off-chip.

The reason for which we chose to assess the potential of the transmittance, reflectance and bounce configurations was the fact that by nature these are the most easily attainable using standard laboratory equipment and circuit/device design techniques. The transmittance configuration requires an optical port of unrestricted (in principle) size or shape and only relatively crude illumination and photodetection

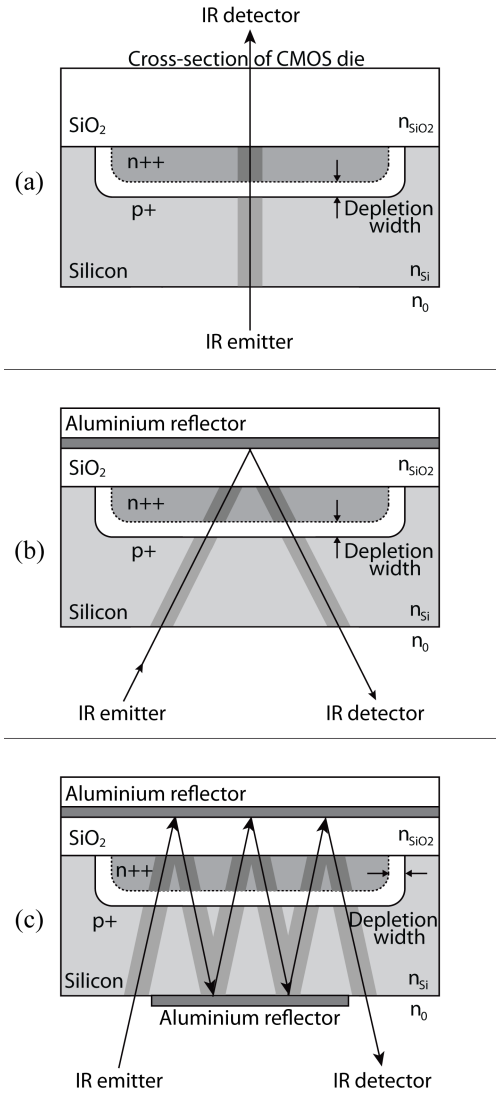

Figure 4.1: Conceptual modulation system indicating the light emitter, a detailed cross-section of the modulator and the photodetector in the three basic configurations. (a): Transmittance configuration. (b): Reflectance configuration. (c): Bounce configuration. Shaded areas along the beam path indicate regions of significant optical losses.

techniques. The reflectance mode adds the constraint that the light source and photodetector must be prevented from communicating directly amongst themselves. Finally, the bounce mode adds the extra complexity of a bottom-side reflector. Yet on balance, all these challenges seemed reasonably solvable. There may be other possible techniques such as shining a light beam in a direction parallel to the surface of the die and using die-wide devices in order to modulate along an extremely long segment of the beam's optical path (much akin to waveguide-based modulation), but these require far better control of illumination conditions and very special modulator designs. Thus the topic of possible geometric configurations for our electro-optical modulation concept is by no means exhausted in this thesis.

The simplest configuration is the 'transmittance' system. We begin by considering the following set-up: A photoemitter, a target die (the actual modulator) lying above the photoemitter facing away from it and a photodetector lying above the target die, facing the photoemitter. Thus the beam is generated on one side of the target die, propagates (is 'transmitted') throughout the entire die, whereupon it undergoes modulation and then exits the other side, ready for detection, hence the name 'transmittance' mode.

Next, in the 'reflectance configuration the photodetector will be lying next to the photoemitter instead of being positioned on the other side of the target die, and will be facing towards the target die. In this configuration too, the beam is emitted towards the target die. It then crosses the entire die, undergoes modulation, hits the reflector integrated within the back-end of the die, above all active components, and then is reflected back downwards. During the descending journey the beam will undergo modulation once more before exiting the die from the underside and carrying on towards the photodetector. This configuration allows a x2 improvement in fractional modulation depth compared to the simple transmittance mode.

Finally, in the 'bounce' configuration everything is positioned as in the reflectance configuration, with the differences that the photodetector is positioned farther away from the light emitter and a reflector material is added at the underside of the die between the near edges of both photoemitter and photodetector (the precise determination of these near edges involves a bit more detail that will be discussed in the appropriate section). Thus, the beam will be emitted towards the target die, and then suffer total internal reflection just like in the reflectance mode case. However, when the 2x modulated, reflected beam reaches the underside of the die, it will once more suffer total internal reflection and start yet another ascending journey towards the surface of the die where it will be once again reflected and start another descending leg. By the time the beam reaches the underside of the die for the 2nd time, it has been modulated 4 times and similarly by the time it reaches the photodetector, after N 'zig-zag' segments, it has been modulated 2N times.

Implementing increasingly complicated system configuration might seem counter-intuitive, but can be beneficial. Let us assume that under certain bias conditions the transmittance through each pass through the die is given by k and under a different set of bias conditions by l. The fractional modulation depth is given by:

$$\frac{\Delta T}{T} = \frac{I_0(k-l)}{I_0 l} = \frac{k}{l} - 1 \tag{4.2}$$

for some initial irradiance I_0 and arbitrarily choosing the state where the losses are l as baseline.

For N passes, the expression becomes:

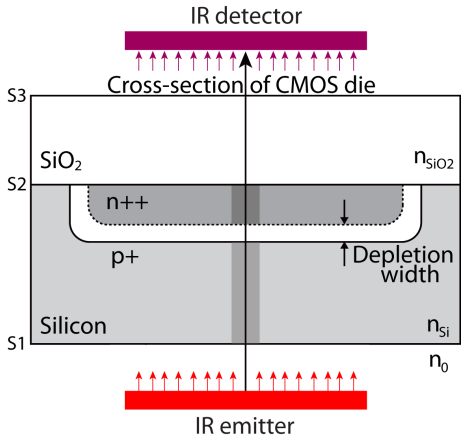

Figure 4.2: Transmittance mode system. An IR emitter sends collimated, uniformly distributed light perpendicular to the die (red arrows) and some of that light is received by the IR detector (magenta arrows). A typical trajectory of a transmitted beam is shown as a large, black arrow. Reflection effects are not shown in this diagram, but will occur at the $Air - Si$, $Si - SiO_2$ and $SiO_2 - Air$ interfaces (dubbed S1-S3 respectively).

$$\frac{\Delta T}{T} = \frac{I_0(k^N - l^N)}{I_0 l^N} = \left(\frac{k}{l}\right)^N - 1 \tag{4.3}$$

This result shows that fractional modulation depth increases with the number of passes through the Silicon. Similar results are obtained regardless of which case is taken as baseline.

In order to study the operation of this device in every configuration, the trajectory of a beam fascicle leaving a point emitter within a narrowly defined angle or the trajectory of a thin fascicle of parallel rays leaving a collimated light source will be followed from emitter to detector. We will call such analysis 'beam fascicle analysis'. In a realistic set-up, an outgoing beam of arbitrary fan-out emanating from a point light source (or rather an approximate equivalent) can be split into different narrow-angle fascicles, each of which follows a slightly different trajectory through the set-up. Similarly a broad, collimated light beam can be split into small, parallel fascicles. Therefore analysis of broad-angle or broad-diameter beams can be performed by simply summing up contributions from individual fascicles within the beams.

The geometric set-up of every analysis revolves around a 2D coordinate system whereby the direction perpendicular to the surface of the target die will be termed the 'vertical' or '0^o' direction against which all angles are measured, whilst the direction parallel to the die surface will be dubbed as the 'horizontal' or '90^o' direction. The substrate is assumed to be p-type and the n-side of the target pn-junction can be of an arbitrary doping distribution.

4.1.1.1 Transmittance mode

Geometrically, the transmittance mode will be considered to revolve around a collimated, uniform light source and a flat detector lying parallel to the surface of the modulator die (see Figure 4.2) The emitted light is assumed to propagate perpendicular to the surface of the die.

We observe that the path from emitter to detector involves three refractive surfaces: $Air - Si$, $Si - SiO_2$ and $SiO_2 - Air$. At each refractive surface the Fresnel relations for reflection and transmission at 0 angle apply (beam falling perpendicularly to the refractive surface)[1]:

$$R_0 = 1 - T_0 = \left(\frac{\eta_s - \eta_t}{\eta_s + \eta_t}\right)^2 \tag{4.4}$$

where R_0 is the reflection coefficient, T_0 is the transmission coefficient, η_s is the 'source side' refractive index (the index in the half-plane from which the light beam arrives) and η_t is the 'target side' refractive index (the index in the half-plane after refraction). Hence forth, whenever an $X - Y$ interface is mentioned, the X side is assumed to be the source side and the Y side the target side. In the 0-angle case, however, because of the square in formula 4.4, transmission and reflection coefficients are exactly the same regardless of the direction of wave propagation.

We can compute reflectance and transmittance coefficients for every refractive interface in advance for given refractive indices, and therefore also implicitly for given wavelengths (see Table 4.1). The wavelengths of interest here lie between $1.3\,\mu m$ and $1.8\,\mu m$.

Table 4.1: 0^o reflection (R_0) and transmission (T_0) coefficients at various interfaces relevant to CMOS technology. Despite the fact that the materials involved are dubbed 'source-' or 'target-' side materials for 0^o angles of incidence transmission and reflection coefficients are exactly the same regardless of wave propagation direction.

Interface	**Source** η	**Target** η	T_0	R_0
$Air - Si$:	1	3.5	0.69	0.31
$Si - SiO_2$:	3.5	1.5	0.84	0.16
$SiO_2 - Air$:	1.5	1	0.96	0.04

Every fascicle that reaches the photodetector will have crossed three media along its path from emitter to detector: air, Silicon and Silicon dioxide. Air causes practically no losses to NIR (Near Infra-Red) beams, as does Silicon dioxide [1]. On the other hand, in Silicon there will be free-carrier absorption losses, which can be described as a sum of Beer-Lambert losses over infinitesimally small slices of lossy material.

The thin-slice Beer-Lambert relation can be derived from the original Beer-Lamber law as such:

$$I(x + \Delta x) = I(x)e^{-\alpha \Delta x} \tag{4.5}$$

where $I(x)$ is light intensity at location x, α is the absorption coefficient and Δx is the slice thickness.

Thus, in order to determine absorptive losses in any lossy medium we need to know the trajectory of the beam through the medium and the absorption coefficient as a function of location along the trajectory of the beam. In our system we have exactly two lossy regions: the lightly (and we assume uniformly) doped p-substrate and the heavily and non-uniformly doped n-diffusion region.

In the substrate region we have a classical case of thick-slice Beer-Lambert absorption due to the FCA phenomenon. Uniform doping means that the absorption coefficient is independent of location within

[1]The following relations apply for both transverse-electric (TE) and transverse-magnetic (TM) waves, unaltered. It is only at non-right angles that TM and TE waves start behaving differently.

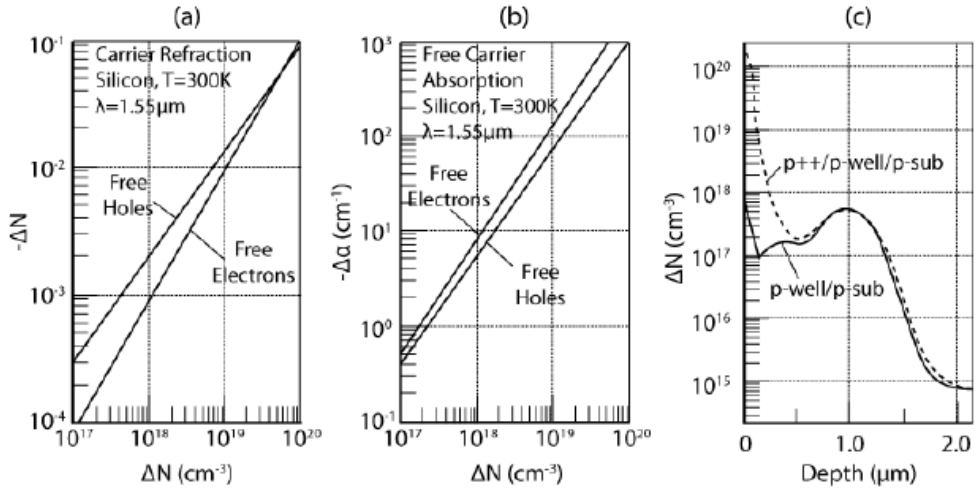

Figure 4.3: The three plots show: (a) Effect of free carrier refraction on the refractive index of silicon as a function of free electrons and free holes. (b) Effect of free carrier absorption on the absorption coefficient as a function of free electrons and holes. (c) Net doping concentration for typical 0.18μm CMOS process. p-type dopant concentrations only are shown for diffusion and implantation. Note: the panel showing free carrier refraction is shown for completeness. Free carrier refraction does not play a role in the mechanism of operation of the communications platform described in this thesis. Panels (a) and (b) are adapted from [3] while panel (c) is reproduced from the transfer report for this project.

the bounds of the lossy region and therefore we may replace Δx with the total path length through the substrate. The said path length will simply be the thickness of the die with the thickness of the diffusion region (up to the metallurgical point) and the thickness of the substrate-side of the depletion region subtracted:

$$L_0 = D - d_{n+} - W_p(V_{bias}) \tag{4.6}$$

where L_0 is the path length through substrate at 0^o angle of incidence, D is the thickness of the die, d_n is the thickness of the n-type diffusion region and W_p is the thickness of the depletion region on the p- (substrate) side of the junction. W_p is determined using theory from section 3.2 and, crucially, is a function of bias voltage.

The value of the absorption coefficient is nominally given by equation 4.1 on page 107, from Schroeder [2] (for the case of free electrons as carriers - the expression for free holes is similar) which can be seen in difference form in the theoretical background chapter (section 3.1.2). Practically, Soref [3] cites that there is a quantitative discrepancy between theory and measurements. However, measured results relating the absorption coefficient to the concentration of free electrons and free holes have already been provided for ranges of interest within the context of CMOS electronics by Soref [3]. An adaptation of these plots can be seen in Figure 4.3.

Plugging the experimentally observed value for the absorption coefficient and the computed value of the path length into the Beer-Lambert relation we can obtain a value K_{sub} for the fractional losses that any beam (of given wavelength) sustains for every perpendicular pass through the substrate:

$$\boxed{I_{end}/I_{start} = K_{sub} = e^{-\alpha_{sub}L_0}} \tag{4.7}$$

where I_{start} and I_{end} are the beam optical intensities (amplitudes) before entering and after crossing the lossy medium (here substrate) respectively and α_{sub} is the absorption coefficient measured for the substrate doping concentration. K_{sub} is the substrate fractional optical loss coefficient.

For the heavily doped diffusion region the situation requires more detailed analysis due to the non-uniform nature of the doping profile. If we assume that the beams pass through the n-diffusion region only far away from the edges and thus edge effects on doping concentration can be neglected, doping concentration becomes only a function of depth (i.e. distance from the surface of the die). We can, therefore, split the path through n-diffusion into infinitesimally thin slices and apply the Beer-Lambert relation in sequence; slice after slice. The optical losses along each infinitesimally small portion of the path throughout n-type Silicon are given by:

$$\frac{dI(x_n)}{dx_n} = -\alpha(x_n)I(x_n) \Leftrightarrow \frac{\delta I(x_n)}{I(x_n)} = -\alpha(x_n) \tag{4.8}$$

where we denoted the location through n-diffusion by means of the variable x_n and $\delta I(x_n) = \frac{dI(x_n)}{dx_n}$.

This equation gives an expression for the optical amplitude 'fractional loss rate' as the beam propagates through the lossy medium as a function of location. The beam intensity upon exiting the lossy medium is hence obtained by means of integrating the fractional loss rate along the entire path from the edge of the depletion region to the $Si - SiO_2$ interface. Following from 4.8:

$$I_{end} = I_{start} \exp\left(-\int_0^{L_{N0}} \alpha(x_n)dx_n\right) \Leftrightarrow \frac{I_{end}}{I_{start}} = \exp\left(-\int_0^{L_{N0}} \alpha(x_n)dx_n\right) = K_0 \tag{4.9}$$

where L_{N0} denotes the total path length through n-type Silicon. The limits of integration are chosen to correspond to setting the surface of the die to 0 (lower limit of integration) and the thickness of the lossy medium (upper limit of integration).

Crucially, L_{N0} is given by the difference between the inherent diffusion region width (up to the metallurgical point) and the depletion region width in the n-side of the junction:

$$L_{N0} = d_{n+} - W_n(V_{bias}) \tag{4.10}$$

The right side version of 4.9 shows that the total loss throughout the n-side of the pn-junction can be expressed as a single fractional loss coefficient K_0 and therefore we can write:

$$\boxed{I_{end}/I_{start} = K_0} \tag{4.11}$$

Because of the summing nature of the integral, this result holds for beams crossing the diffusion region in either direction so K_0 truly represents the fractional loss per pass through the diffusion region

at 0^o angle of incidence. Crucially, because W_n is also a function of voltage bias (just like W_p earlier) K_0 will too be a function of bias voltage.

Finally, we observe that because of the lack of a refractive surface between the substrate and the diffusion region (refractive index perturbation due to differences in free carrier concentration are assumed to be negligibly small -see Figure 4.3, panel (a)-), any beam passing through one lossy medium will also pass through the other. This immediately implies that we can aggregate losses in the p-substrate and the n-diffusion region in order to obtain an aggregate Silicon fractional loss coefficient K_{Si}:

$$\boxed{K_{Si} = K_{sub}K_0 = e^{-\alpha_{sub}L_0}K_0} \tag{4.12}$$

Now we have all the background to start the analysis of a beam fascicle trajectory in earnest. The beam fascicle analysis proper begins by denoting the initial amplitude of the generic fascicle under study as I_0. The beam that starts with an amplitude of I_0 can then be split into a quasi-infinite number of components that bounce back and forth between refractive surfaces different numbers of times. We assume that any radiation crossing the $Air - Si$ interface (S1) towards the photoemitter is lost and any crossing the $SiO_2 - Air$ interface (S3) is detected, which is why we are particularly interested in components that have been subjected to refraction even numbers of times (and therefore end up exiting the die upwards via S3). As such, the notion of a 'reflection pair' will play a crucial role in the following analysis. A reflection pair is simply the pairing of a reflection that causes an ascending beam to switch direction and start heading downwards followed by a reflection that diverts the beam back to its original upward direction.

Notable components of the beam would include the main component that has been transmitted through each refractive surface. The total transmitted power of the main component is simply I_0 multiplied by the transmittance coefficients of each refractive interface and the coefficient that describes losses in Silicon:

$$I_{NULL} = I_0T_1K_{Si}T_2T_3 \tag{4.13}$$

where I_{NULL} represents the component that has not undergone any refraction at any surface and T_x represents the transmission coefficient for surface x. This component will be the only member of what we will call 'set G_0', the set of all beam fascicles that have not undergone any reflections (the subscript indicates the number of reflections suffered by the beam components that belong to the set).

The next group of components (set G_2) contains all components that suffer a single pair of reflections, and therefore also reach the photodetector eventually:

$$I_A = I_0T_1K_{Si}R_2K_{Si}R_1K_{Si}T_2T_3 = I_{NULL}(R_2K_{Si}R_1K_{Si}) = I_{NULL}M_A \tag{4.14}$$

$$I_B = I_0T_1K_{Si}T_2R_3R_2T_3 = I_{NULL}(R_3R_2) = I_{NULL}M_B \tag{4.15}$$

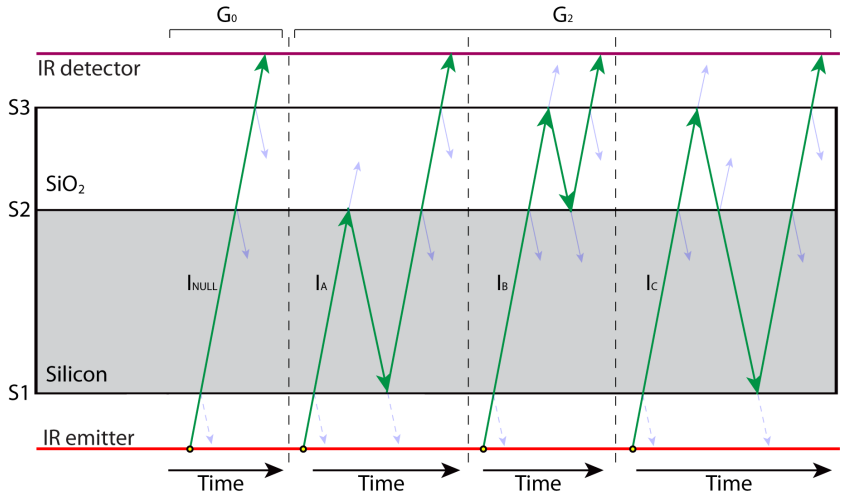

Figure 4.4: Important components of the emitted beam as described by equations 4.13-4.16. The green arrows indicate the path of the beam component under study, the pale blue arrows indicate the paths of other beam components. Dashed pale blue arrows indicate paths of components that are permanently lost. The yellow/black dot at the start of each beam component's path denotes time t_0. The internal structure of the Silicon region of the die and differences in the speed of light due to different refractive indices are not illustrated for clarity.

$$I_C = I_0 T_1 K_{Si} T_2 R_3 T_2 K_{Si} R_1 K_{Si} T_2 T_3 = I_{NULL}(R_3 T_2 K_{Si} R_1 K_{Si} T_2) = I_{NULL} M_C \tag{4.16}$$

where the subscript in I_x means:

- A: The pair of reflections occurs at surfaces S2 and S1, i.e. the beam bounces once within the Silicon.
- B: The pair of reflections occurs at surfaces S3 and S2, i.e. the beam bounces once within the Silicon dioxide.
- C: The pair of reflections occurs at surfaces S3 and S1, i.e. the beam bounces once throughout the entire width of the die.

As an example I_A is transmitted through surface 1, reflected at surface 2, reflected at surface 1, transmitted through surface 2 and transmitted through surface 3; a process that is reflected in each term of the fully expanded product of terms in equation 4.14. A K_{Si} is added every time the beam traverses the space between S1 and S2 as optical losses will occur for each such traversal. The two rightmost sides of equations 4.14-4.16 highlight the terms introduced by the reflective bounce in each case and dub them as M_x, where x denotes the type of reflection pair. Clearly, the members of set G_2 consist of the sole member of G_0 multiplied by a constant smaller than unity that describes the effect of the introduced reflection pair. Members of sets G_0 and G_2 are shown in Figure 4.4.

Set G_4, whose members are subject to two pairs of refractions, will be constituted of elements formed by the basic I_{NULL} factor multiplied by one of the possible combinations of additional factors M_A, M_B

and M_C. Possible combinations are: M_BM_B, M_BM_C, M_CM_B, M_CM_C, M_AM_C, M_AM_B and M_AM_A. The set of possible combinations is determined by geometry and for this simple case of three refractive/reflective surfaces can be found by exhaustive searching and validity checking[2].

We note that M_BM_C and M_CM_B may well be subject to exactly the same losses, but they still form distinct parts of the beam. Specifically, M_BM_C arises from the part of the beam that is reflected upon hitting surface S2 for the second time (during a descending leg), whilst M_CM_B arises from the component that is transmitted upon hitting S2 for the second time. In the general N reflection pair case possible combinations of M factors containing the same M_x components in different order are separate factors and need to be considered individually. However, in the case where more than one factor of some type M_{x0} exists within the transmission expression it only needs to be considered once for all permutations of all M_{x0} factors amongst themselves. In other words swapping two factors M_x and M_y with each other within a string of M-factors creates two terms that refer to different beam components and therefore need to be taken into account separately if $x \neq y$ but refers to the same beam component and need only be taken once if $x = y$. Thus, for example $M_A^2M_C$ only needs to be considered once whilst $M_A^2M_CM_B$ and $M_AM_BM_CM_A$ are distinct components, assuming they refer to a set-up configuration where they are all valid combinations (M_B swapped place with one of the M_As from the M_A^2 term.).

Finding a suitable mathematical expression for the amount of light reaching a photodetecor detector including all components has been previously done in transfer matrix form [4] and can yield exact solutions so long as all parameters are set to accurately represent the optical media. However, in search of a simpler and 'reasonably good', rough solution we decided to sum the first few components. This will yield a lower bound as to how much radiation reaches the photodetector. This analysis needs to be done numerically with appropriate values for K_{Si}. In the simple example of taking into account nothing but the main component I_{NULL} we notice that the aforementioned main component of the beam carries approximately $(55.6\%)K_{Si}$ of the beam amplitude directly to the photodetector, whilst 31% is permanently lost at the $Air{-}Si$ refractive surface (S1). This yields the very crude $(55.6\%)K_{Si}{-}79\%$ lower and upper bounds for beam energy transmitted through an IR optical link at 1.3-1.8 micron wavelength perpendicularly to the die in the transmittance system configuration. Taking more components into consideration will reduce the interval between upper and lower bounds and gradually converge towards the correct answer.

The final piece of the puzzle involves computing values for the optical loss through Silicon factor K_{Si} for different bias voltages. As we have seen in this chapter the bias voltage influences the width of the depletion region, which then influences how wide the regions of significant optical losses within Silicon will be and consequently changes the value of K_{Si}. Once total transmitted beam power is computed for K_{Si} factors corresponding to both desired bias voltage levels the difference in transmitted optical power per fascicle can be extracted and multiplied by the horizontal extent of the set-up (in units of fascicle widths) in order to yield the difference in detected light intensity:

$$\boxed{\Delta I_{det} = X|I_{det}(K_{Si}(V_{low})) - I_{det}(K_{Si}(V_{high}))|} \tag{4.17}$$

where I_{det} is the numerically extracted light intensity that reaches the photodetector and is shown explicitly as a function of K_{Si}, which is in turn explicitly shown as a function of voltage bias and X is

[2]Rules for finding possible combinations of N reflection pairs can be found for the general case of Q refractive surfaces, but determining them is outside the scope of this thesis.

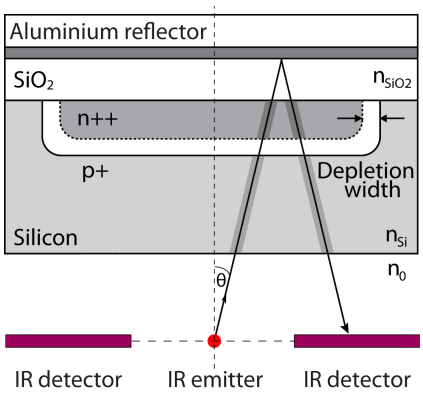

Figure 4.5: Reflectance mode system. The IR emitter is a point light source and the detector consists of two segments parallel to the die surface and coplanar with the light source. This set-up is used for symmetry although most of the analysis concerns parametric examination of a single beam fascicle, just like the one shown in the figure as a black path between emitter and detector. The so-called 'angle of departure' that describes the angle between a beam fascicle and the direction normal to the die surface and is symbolised by θ is also shown.

the horizontal extent of the emitter and the detector.

Equation 4.17 shows that it is possible to modulate the amount of light received by a photodetector by tampering with the bias voltage applied across the modulating die. If this change in light intensity is within the detectability limits of the detector and associated instrumentation, then transmitting information from a die 'contactlessly' via an optical link in transmittance configuration is possible.

Generalising this set-up to the 3D case involves the trivial operation of multiplying the 'fascicles', which, as is abundantly clear by now represent nothing more than energy density in both 2D ad 3D cases, by the area of the system A.

4.1.1.2 Reflectance mode

Geometrically the reflectance mode system revolves around a point light source and a 2-segment photodetector lying parallel to the surface of the modulator die on the same plane as the photoemitter (see Figure 4.5). The analysis in this section is an elaboration on our previous publication [5].

The beam fascicle analysis begins by denoting the initial amplitude of the generic fascicle under study as I_0 and the angle at which it departs from the light source with respect to the direction normal to the die surfaces (the 'angle of departure') as θ. The fascicle crosses the air-gap separating the emitter and the die with negligible losses. At the air-Silicon interface some of the light will be transmitted and some will be reflected. The generic angle Fresnel relations describing EM wave reflection at a heterogeneous surface[3] interface are:

$$R_m = 1 - T_m = \left(\frac{\eta_s \cos(\theta) - \eta_t \sqrt{1 - (\frac{\eta_s}{\eta_t} \sin(\theta))^2}}{\eta_s \cos(\theta) + \eta_t \sqrt{1 - (\frac{\eta_s}{\eta_t} \sin(\theta))^2}} \right)^2 \tag{4.18}$$

[3]I.e. where two materials of different refractive indices meet.

$$R_e = 1 - T_e = \left(\frac{\eta_s \sqrt{1 - (\frac{\eta_s}{\eta_t} \sin(\theta))^2} - \eta_t \cos(\theta)}{\eta_s \sqrt{1 - (\frac{\eta_s}{\eta_t} \sin(\theta))^2} + \eta_t \cos(\theta)} \right)^2 \tag{4.19}$$

where (4.18) refers to the reflection coefficient in the TM (transverse magnetic) mode, and (4.19) to reflection in the TE (transverse electric) mode. R_m is the TM reflection coefficient, R_e is the TE reflection coefficient, θ is the angle of incidence of the beam (equal to the angle of departure since the refractive surface is perpendicular to the vertical direction), η_s is the source-side refractive index (air) and η_t the target-side refractive index (Silicon). The corresponding transmission coefficients for TE and TM modes respectively are marked as T_e and T_m respectively.

If we assume that the emitted light is unpolarised, i.e. that light is distributed evenly across all possible polarisation angles, then the intensity of the beam remaining after crossing the air-Si interface (I_1) is given by:

$$I_1 = I_0 \left(\frac{T_m + T_e}{2} \right) \tag{4.20}$$

since decomposing the incident beam into TE and TM components we find that given our assumptions their contributions to the fascicle energy are equal.

The direction of the beam is also altered by crossing from the low refractive index air-gap to the high refractive index Silicon in accordance to Snell's law:

$$\eta_s sin(\theta) = \eta_t sin(\theta_1) \tag{4.21}$$

where θ_1 represents the angle at which the beam travels within the Silicon bulk.

The next segment of the beam's path is through the Silicon bulk, from the air-Si interface to the target pn-junction's depletion region edge. If we assume that the doping concentration from the bottom of the die up to the edge of the depletion region is constant, then the losses simply follow the Beer-Lambert relation described in section 3.1.1, although the physical phenomenon causing Beer-Lambert decay is not inter-band absorption, but rather free carrier absorption. The Beer-Lambert relation is reproduced with appropriate adaptations here for the reader's convenience:

$$\boxed{I_2 = I_1 e^{-\alpha L} = I_1 K_{sub}} \tag{4.22}$$

where I_2 is the intensity of light remaining in the fascicle by the time the radiation reaches the edge of the depletion region of the target pn-junction and L represents the path length from the air-Si interface to the depletion region edge of the target pn-junction. In order to determine L we use trigonometry:

$$L = \frac{L_0}{cos(\theta_1)} \tag{4.23}$$

where L_0 is the minimum distance between the air-Si interface and the edge of the depletion region of the target pn-junction (i.e. the 'perpendicular distance') and is given by 4.6, reproduced here for convenience:

$$L_0 = D - d_{n+} - W_p(V_{bias}) \tag{4.24}$$

The absorption coefficient, on the other hand, is dependent on the doping concentration profile throughout the trajectory of the beam. Nominally, this will be given by equation 4.1 on page 107 (for free electrons as carriers).

Practically measured results of α as a function of free carrier concentration can be taken by use of the results seen in Figure 4.3.

In the next path section the beam crosses depleted Silicon, where there will be virtually no amplitude attenuation. As such the beam intensity upon exiting the depletion region (I_3) will be equal to the intensity upon entering the depletion region:

$$I_3 = I_2 \tag{4.25}$$

The beam then enters a doped region that forms the n-side of the pn-junction. This will be characterised by a certain doping profile in the 'vertical' direction $N_D(x)$. Using 4.1 we find that similarly the absorption coefficient will be a function of location $\alpha(x)$. Correcting for the angle by using equation 4.23 we obtain:

$$\alpha(l) = \alpha(x/cos(\theta_1)) \propto N_D(x/cos(\theta_1)) \tag{4.26}$$

where l denotes distance traveled within the n-type material expressed in rotated coordinates (rotated y-axis is aligned parallel to the beam).

Using the same methodology that was used for the transmittance case (equations 4.8 and 4.9), but this time for a rotated set of coordinates we obtain an integral that describes losses through the diffusion region for a beam fascicle crossing it at a generic angle:

$$I_4 = I_3 \exp\left(-\int_0^{L_N} \alpha(l)dl\right) \Leftrightarrow \frac{I_4}{I_3} = \exp\left(-\int_0^{L_N} \alpha(l)dl\right) = K \tag{4.27}$$

where the integral is taken along the beam path through non-depleted n-diffusion type Silicon and L_N represents the total path length through n-type Silicon and is given by:

$$L_N = \frac{d_{n+} - W_n(V_bias)}{cos(\theta_1)} \tag{4.28}$$

The right side version of 4.27 shows that the total loss throughout the n-side of the pn-junction can be expressed as a single fractional loss coefficient K and therefore we can write:

$$\boxed{I_4 = KI_3} \tag{4.29}$$

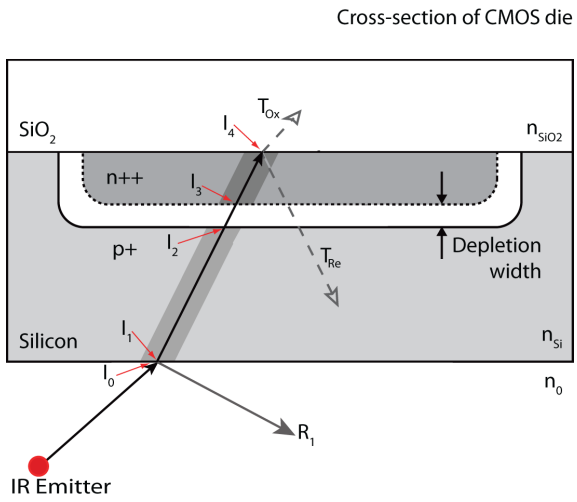

Figure 4.6: Modulator cross-section with the ascending path shown alone for clarity (in black). R_1 represents reflected light that is assumed to be lost for all configurations. Points marked as I_x where x is an index number refer to the corresponding points from equations 4.20 to 4.29. Beyond the $Si - SiO_2$ interface, I_4, the beam splits into a transmitted component that carries on through the Silicon dioxide (T_{Ox}) and a reflected component that begins a descending journey (T_{Re}). The significance of each of these components will depend on the configuration of the system and the angle of departure.

where I_4 is the beam intensity upon hitting the $Si - SiO_2$ interface.

The ascending leg of any beam fascicle trajectory for a generic angle is shown in Figure 4.6.

Once the beam hits the $Si - SiO_2$ interface, the Fresnel relations describing absorption and refraction would normally need to be employed once again. However, the transmitted part of the beam (T_{Ox}) will continue upwards for a minute distance (typically in the 1 micron range or less) before suffering total internal reflection at the $Al - SiO_2$ interface and starting the downward leg of the journey. By the time it hits the $Si - SiO_2$ interface on its descending leg we can consider that the lateral displacement of T_{Ox} wth respect to the previously reflected part of the fascicle T_{Re} will be negligible. As such, both T_{Ox} and T_{Re} can be generally assumed to eventually hit the photodetecting target.

Matters, however, are complicated by the fact that T_{Ox} will encounter the $Si - SiO_2$ interface once again on the downward leg of its journey. The Fresnel relations can once more be employed to calculate how much of it will be reflected upwards and towards the Aluminium reflector and how much of it will continue the downward leg. The reflected part will once again meet the Aluminium reflector and begin descending on yet another downward leg. This 'ping-pong' process will be iterated for relatively few times before most of the beam energy leaks into the Silicon, and therefore most of the T_{Ox} energy can be assumed to eventually reach the photodetecting target. This applies so long as the angle of incidence at the $Si - SiO_2$ interface is not too close to or above the critical angle[4]. In that case the beam will either exit the SiO_2 trap having leaked most of its energy far away from the photodetecting target, or only at the edge of the chip. This case will be discounted since the system is not designed to operate with light emitted at angles that would cause 'ping-ponging' to dramatically affect operation. Moreover, the effect is likely to be of any significance only for a very narrow band of angles, although this is a matter for further study.

[4] SiO_2 has a lower refractive index than Silicon in the wavelength span of interest (1.3-1.8 μm).

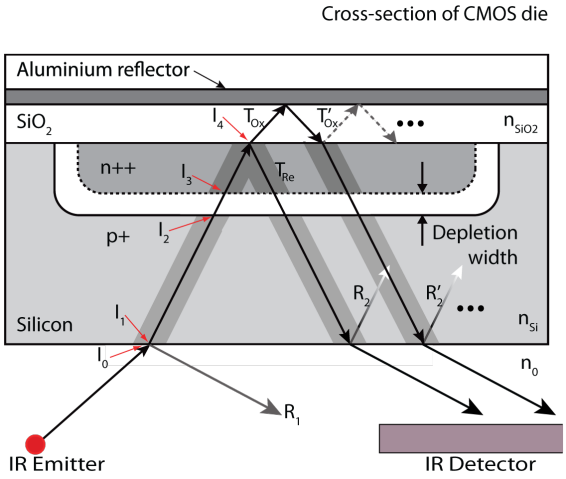

Figure 4.7: Full modulation system indicating the light emitter, a detailed cross-section of the modulator and the photodetector. The ascending beam along with the main descending beams are shown. Points $I_x| \in \{0, 1, 2, 3, 4\}$ correspond to the intensities of the beam at various points along the trajectory as described by equations 4.18 - 4.30. T_{Re} labels the part of the beam reflected at the $Si - SiO_2$ interface, while T_{Ox} represents the part of the beam that continues through the oxide. $R_x|x \in \{1, 2\}$ marks reflective losses that are assumed to never reach the photodetector. Finally, T'_{Ox} shows the part of the beam that is subjected to the 'ping-pong' effect.

Furthermore, since SiO_2 is a material that offers very low optical resistance to near infra-red (NIR) beams [1], we can assume that T_{Ox} and any beam fascicles suffering 'ping-pong' effect are not exhibiting any significant extra losses compared to T_{Re} by the time they re-enter Silicon territory[5].

Thus we reach the conclusion that for the purposes of analysis we can treat the multiple, parallel descending beams as a single descending beam fascicle of intensity I_4. This is illustrated in Figure 4.7.

The analysis for the descending leg of the beam is very similar to that for the ascending leg, particularly since the symmetry of Snell's law implies that the beam will cross each segment of the descending path at the same absolute angle as has occurred for the corresponding ascending part of the path (i.e. the path will be symmetrical about a line parallel to the vertical direction that crosses through the point at which the beam is totally internally reflected by the Aluminium reflector). Therefore, the losses through n-type and bulk Silicon will be exactly the same as for the ascending leg, but once the beam reaches the Si-air interface the Fresnel relations will need to be applied once more. Since the refractive index of Silicon is higher than that of air, reflection will take a toll on the amount of light exiting the Silicon substrate. This time internally reflected light is assumed to be lost because of the large lateral displacement of any 'ping-ponged' beams that will be created by the said totally internally reflected beam at the surface of the die. Total internal reflection can also be an issue for beams hitting the Si-air interface at high angles (discounted here).

Finally, the part of the beam that has successfully exited Silicon and hits the detector is sensed and has suffered a total loss along the path from emitter to detector that is described by:

[5] At least for a 'reasonably' low number of 'ping-pong' iterations.

$$\boxed{I_{det} = \gamma I_0 \,|\, \gamma = \left(\frac{T_m + T_e}{2}\right)_{asc} \exp\left(-2\alpha \frac{D}{cos\theta_1}\right) K^2 \left(\frac{T_m + T_e}{2}\right)_{desc}} \tag{4.30}$$

where I_{det} is the detected beam intensity, γ is the optical loss factor throughout the entire path length the *asc* and *desc* subscripts denote elements relating to ascending or descending legs respectively. Fractional losses via the free-carrier absorption are equal on the ascending and descending legs, as is the reflection-refraction balance for crossing the $Si - Air$ interface in either direction. This occurs because of the inherent symmetry of the Fresnel relations when used in combination with Snell's law. In other words by interchanging the refractive indices in equations 4.18 and 4.19 and also replacing the angle of air to Si incidence with the angle of air to Si refraction (= angle of Si to air incidence due to symmetry in our set-up) as dictated by Snell's law we obtain the same reflection/refraction balance, namely transmission and reflection coefficients, in both cases. The terms describing this balance for ascending and descending legs have been kept separate despite being numerically equal in order to stress the fact that they refer to different angles of incidence and different source and target refractive indices.

Equation 4.30 describes the optical losses at a given steady-state, i.e. where the bias across the pn-junction is held constant, as a function of angle of refraction at the air-Si interface, distance from the bottom of the die to the edge of the depletion region of the target pn-junction and the lumped K factor. In a modulator the value of I_{det} will need to be changed somehow between at least a couple of distinct states (ON and OFF). Thus, if we consider the angle of departure θ a parameter, the only lever remaining to the engineer is modulation of the width of the depletion region; a task that can be achieved by appropriate changes in the biasing of the pn-junction.

In practical terms this is expressed as a change in the value of L_0, as well as a shift in the limits of integration hidden within K. Let us denote the 0-bias perpendicular distance between the bottom of the die and the edge of the depletion region as $L_{0,0}$, and the 0-bias limits of integration for K as 0 and L_N. If we now apply a reverse bias to the pn-junction, the depletion region will widen, thus creating the reverse-bias perpendicular distance between chip underside and depletion region edge L_r and shifts the limits of integration of K to δx and L_N:

$$K_r = -\exp\left(\int_{\delta x}^{L_N} \alpha(l) dl\right) \tag{4.31}$$

where K_r represents the reverse bias value for K.

In order to calculate the change in L_0 and the value of δx as functions of applied reverse bias voltage we need to revisit section 3.2 and also create a concrete doping profile model for the n-side of the pn-junction (perhaps with help from section 3.3). This will ultimately yield an expression for the absolute difference in optical losses between the full-ON and full-OFF states ΔI_{det}:

$$\boxed{\Delta I_{det} = I_0 \left(\frac{T_m + T_e}{2}\right)_{asc} \left(\frac{T_m + T_e}{2}\right)_{desc} \left[\exp\left(-2\alpha \frac{L_0}{cos\theta_1}\right) K_0^2 - \exp\left(-2\alpha \frac{L_r}{cos\theta_1}\right) K_r^2\right]} \tag{4.32}$$

where K_0 represents the value of the K factor for 0-bias.

Generalising the concept from a beam fascicle to a wide fan-out beam

In order to examine the case of a wide fan-out beam we shall make some simplifications and specify the geometry of the set-up anew, in accordance with the simplifications. Thus:

- The 'ping-pong' effect shall be completely ignored. This is acceptable given the low optical losses through SiO_2 for the wavelengths of interest and the small extent of beam lateral displacement that it causes for low angles. In the case of high angles where the ping-pong effect cannot be ignored it will be optical losses at the Si-air interface that will dominate long before the $Si - SiO_2$ refraction/reflection relationship becomes significant. This stems from the fact that both interfaces involve Silicon, but air has a significantly lower refractive index than SiO_2.
- The SiO_2 slice will be completely ignored. This follows from the previous point and the fact that the thickness of the SiO_2 layer is very small in comparison to the thickness of the die.
- The refractive index of Silicon does not change significantly for the doping concentrations involved. This is a reasonable approximation given the results presented in [6], whereby for doping and free carrier concentrations up to at least 10^{18} dopants per cubic cm the change in refractive index vs. undoped Silicon is limited to a maximum of about $2 \times 10^{-3}/cm$. The main results in the reference correspond to wavelengths of 1.3 and 1.6 μm and show little variation as a function of wavelength.

The new geometric set-up will consist of a point light source, a target die of specified thickness with a pn-junction of a specified geometry and a detector of given geometric specs. The set-up is illustrated in Figure 4.8, complete with variables that will be used throughout this section. The variable d will denote the distance between the light source and the underside of the Silicon die, whilst the variable D will denote the thickness of the die from the underside to the surface (reflector layer). The geometry of the set-up is 2D, but we shall generalise it to 3D by assuming that the target pn-junction is a disk and the detector is an annulus.

The origin of the beams, the point light source, will have some arbitrary intensity distribution as a function of angle of departure. This will be denoted as $Z(\theta)$ with units of W/rad in the 2D case and W/sr in the 3D case. Next, the optical loss along a beam path from emitter to a hypothetical detection point will also be a function of angle of departure. This will be given by equation 4.30, but now the units will be rad^{-1} for 2D and sr^{-1} for 3D. Finally, the detector needs to be defined geometrically since only a fraction of the beam (to be determined) will reach the photodetector.

For simplicity let us assume that the photodetector is a single, annular 2D-structure (i.e. thickness = 0) that sits on the same plane as the point light source. The point light source will be the centre of the annulus, and therefore the distance between the annulus and the point light source will be equal to the inner radius of the annulus, a value we will mark by X_{in}. The distance between the point light source and the outer rim of the photodetector will be denoted as X_{out}.

The next step is to determine the range of angles of departure within which outgoing beams are expected to hit the photodetector. Under the assumption that the lateral displacement experienced by the beam during its ascending and descending legs is equal due to symmetry, we obtain:

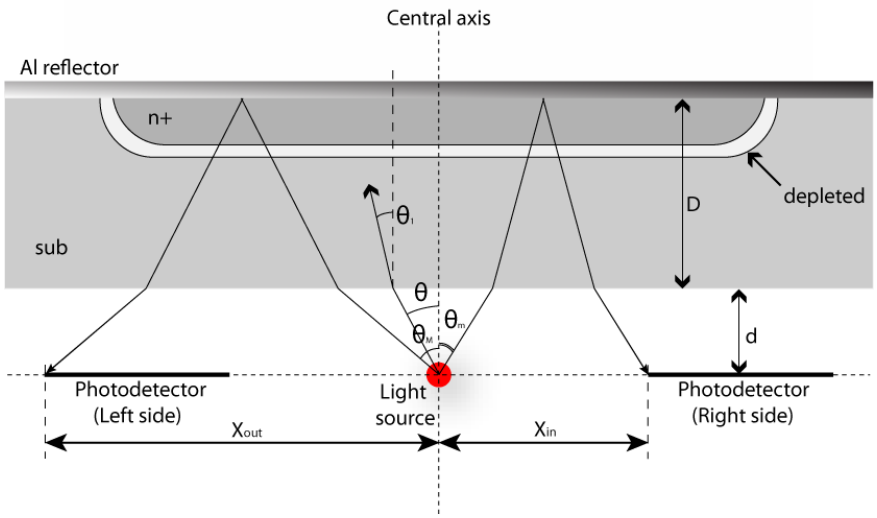

Figure 4.8: Simplified system used throughout this section for deriving the relationship that describes power delivery from the light source to the photodetector via the modulator. d: Distance between light source and the underside of the die. D: Die thickness. θ_m: Minimum critical angle. θ_M: Maximum critical angle. θ: generic angle of departure. θ_1: Angle through Silicon given that angle of departure is equal to θ. n+: n-type diffusion region. sub: p-type substrate. Note: the beam fascicle shown heading leftwards departs the light source at the maximum critical angle whilst the one illustrated heading rightwards departs at the minimum critical angle. This is for reasons of clarity. In reality beam fascicles will depart in every direction.

$$X_{in} = 2[d\tan(\theta) + D\tan(\theta_1)] \tag{4.33}$$

where we can replace θ_1 with a function of θ by use of Snell's law and use basic trigonometry to obtain:

$$X_{in} = 2\left(d\tan(\theta) + D\frac{\frac{n_s}{n_t}\sin(\theta)}{\sqrt{1-\left(\frac{n_s}{n_t}\right)^2\sin^2(\theta)}}\right) \tag{4.34}$$

or rather more elegantly:

$$X_{in} = 2\left(\frac{d\sin(\theta)}{\sqrt{1-\sin^2(\theta)}} + \frac{D\frac{n_s}{n_t}\sin(\theta)}{\sqrt{1-\left(\frac{n_s}{n_t}\right)^2\sin^2(\theta)}}\right) \tag{4.35}$$

Solving equation 4.35 numerically will yield a value close to the 'minimum critical angle' θ_m below which emitted beams miss the photodetector. Replacing X_{in} with x_{out} and solving once again will similarly yield the 'maximum critical angle' θ_M above which the emitted beams will miss the photodetector.

We now have enough information to calculate the power the reaches a well-specified photodetector in the 2D case (P_{2D}):

$$\boxed{P_{2D} = 2\int_{\theta_m}^{\theta_M} [Z(\theta)\gamma(\theta)]^2 \, d\theta} \tag{4.36}$$

where the square within the integral arises because of the need to convert amplitude into power. The factor of 2 is introduced to reflect the fact that in the 2D case there are exactly 2 beams that leave the emitter at an absolute angle of θ (one either side of the axis of symmetry).

This result is generally true for any set-up. The specific geometry of the detector will influence the limits of integration, the geometry of the target die will influence the γ factor and to some extent the limits of integration, the power distribution of the point light source will determine Z and the factor of 2 will apply in setups that are symmetric about the central axis only.

In order to render this result into its 3D equivalent we need to take into account the fact that we need to integrate over solid angles (steradians) as opposed to simple arcs:

$$\boxed{P_{3D} = 2\int_{A} [Z(\Theta)\gamma(\Theta)]^2 \, d\Theta} \tag{4.37}$$

where A is the angle surface covered (determined by geometry), Θ denotes the angular surface element and functions Z and γ are in their 3D versions. This is the general solution for 3D.

In the case of our simple setup (disk target and annular detector) we can exploit symmetry to simplify the equations. At this point we need to make assumptions about the nature of the point light source:

Case 1: The point light source intensity is only a function of θ. We simply need to add a correction term to the power equation. We will call this the 'arc correction' term:

$$R(\theta) = sin(\theta) \tag{4.38}$$

This factor will account for the fact that the total beam energy radiated via circles corresponding to different 'latitudes' (angular elevation) of a sphere of arbitrarily small radius with the point light source at its centre. For example the total energy of all fascicles intersecting the sphere at latitude 80^o will be far smaller than the corresponding energy for all fascicles intersecting the sphere at the equator. Thus the corrected power equation now becomes:

$$P_{3D} = \int_{\theta_m}^{\theta_M} [Z(\theta)\gamma(\theta)R(\theta)]^2 \, d\theta \tag{4.39}$$

where the symmetry of the set-up has removed the need of taking into account azimuthal rotation thereby allowing us to express the problem in terms of just the polar angle θ, just like in the 2D case.

Case 2: Perfectly homogeneous point light source. Equation 4.39 now becomes:

$$P_{3D} = Z^2 \int_{\theta_m}^{\theta_M} [\gamma(\theta)R(\theta)]^2 \, d\theta \tag{4.40}$$

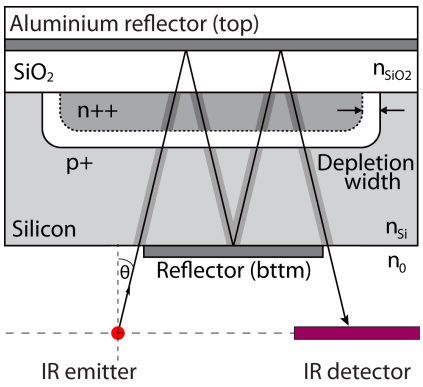

Figure 4.9: Bounce mode system. The IR emitter is a point light source and the detector consists of a segment parallel to the die surface and coplanar with the light source. In the illustrated case the beam departs at such angle as to bounce against the bottom reflector only once along its path from emitter to detector. Angle of departure θ is also shown for this particular fascicle, as is the horizontal displacement between photemitter and near edge of the bottom reflector (W_r) and the vertical displacement between photoemmiter and the top of the bottom-side reflector (H_r).

Case 3: The point light source has a truly arbitrary intensity distribution vs both azimuthal and polar angles. In this case we cannot exploit geometrical symmetry to any significant extent and thereby need to revert to equation 4.37.

4.1.1.3 Bounce mode

In this configuration the system is organised in almost exactly the same way as for reflectance mode, with the difference that between the photoemitter and the photodetector the die is covered by both a top layer and a bottom layer metallic reflector (Figure 4.9). The positioning of the bottom reflector, specifically the horizontal displacement of its near edge (near to the photoemitter) with respect to the photoemitter (W_r) and the vertical distance between the top of the bottom-side reflector and the photoemitter (H_r) determines the departure angle range for which emitted beams are captured by what is in essence a waveguide formed of two metal layers. The detector is assumed to be of infinite length, as is the top Aluminium metallisation, which means that all fascicles entering the waveguide will eventually hit the photodetector.

To begin we take a look at the $Si - SiO_2$ interface, which introduces much complexity in the system. Because Silicon has a higher refractive index than its dioxide counterpart there will be three important ranges of angles of $Si - SiO_2$ incidence:

The first one (R_1) is for acute angles of incidence. In this angle of incidence we get the effect of beam-multifurcation which requires careful consideration (see later). This region extends between 0^o and about 25^o and its defining characteristic is that because of the high ratio of die thickness over Silicon dioxide layer thickness (of the order of 100:1) the lateral displacement of a beam through Silicon will be much larger than that through Silicon dioxide. The precise definition of what is a 'much larger' lateral displacement through Silicon versus oxide will determine the exact upper limit of this region. Specifically horizontal displacement is given by:

$$\Delta x = \tan(\theta) D \tag{4.41}$$

where x denotes location, θ the angle at which the beam traverses the medium and D the thickness of the medium.

The ratio of horizontal displacements between two media of different thicknesses and beam traversing them at different angles is given by:

$$\Gamma = \frac{\tan(\theta_1) D}{\tan(\theta_2) d} \tag{4.42}$$

where D is the thickness of the thicker medium, d that of the thinner medium and θ_x the angles of traversal through each medium. By substituting one of the θ_x angles by use of Snell's law one may determine the ratio of horizontal displacements as a function of geometry and angle of incidence. The resulting relation turns out to be:

$$\Gamma = \frac{D\sqrt{1 - \left(\frac{n_s}{n_t}\right)^2 \sin^2(\theta_1)}}{d\frac{n_s}{n_t}\sqrt{1 - \sin^2(\theta_1)}} \tag{4.43}$$

where n_s is the source-side refractive index and n_t is the target-side refractive index for the wavelengths of interest. Γ has now been expressed without the angle of refraction.

By setting Γ to a certain acceptable value (for example 50, in order to make sure that lateral displacement in the oxide region is much smaller than that in the Silicon) and solving for θ one can set a suitably defined cut-off point for the transition between the first and second angle of incidence regions. In our case plugging in the correct values of the various variables will yield a value of approximately 23^o (highest integer value satisfying $\Gamma \geq 50$), which is close to the roughtly 25^o angle mentioned before.

The second angle of incidence region (R_2) is for a narrow interval of angles of incidence for which the angle of refraction in Silicon dioxide is very close to 90^o. It stretches between the border between R_1 and R_2 and up to the point where the angle of incidence becomes the critical angle of refraction. This region is defined by the horizontal displacement of a beam fascicle when traveling through Silicon from the bottom of the die to the $Si - SiO_2$ surface becoming comparable to the equivalent displacement through the much narrower Silicon dioxide layer. Such region can be proven to exist because as the angle of incidence tends towards the critical refraction angle the displacement through Silicon tends to a finite value while the displacement through Silicon dioxide tends to infinity. For simplicity this region will be discounted. This can also be done given that at angles of incidence so close to the critical angle of refraction most of the beam energy is subjected to reflection rather than refraction.

The third and final range (R_3) is the trivial case where total internal reflection occurs and the Silicon dioxide layer can be ignored completely in every way. These regions are illustrated in Figure 4.10.

An effect that is set to complicate analysis of the bounce configuration and occurs in the acute angle range (R_1) is beam multifurcation: Following from the discussion in the reflectance section (4.1.1.2) the SiO_2 layer will cause our fascicle to multifurcate in the horizontal direction (for visualisation purposes please refer back to Figure 4.7) thus creating a main component that is always transmitted through the $Si - SiO_2$ interface and 'child components' that are subjected to reflections at the single remaining

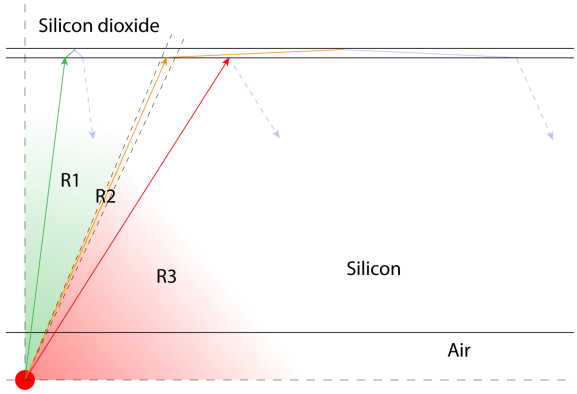

Figure 4.10: The three incidence angle ranges within which the behaviour of incident beams is significantly different from the perspective of a bounce mode system. R1 (green) region: lateral displacement through the dioxide layer is small compared to the corresponding displacement through Silicon. R2 (orange) region: the displacement through the dioxide layer is comparable to that through Silicon. R3 (red) region: total internal reflection at the $Si - SiO_2$ interface. Example beams are also shown for each region. The refractive effects at the $air - Si$ interface have been omitted for clarity although in reality they would have given each beam shown an extra lateral displacement through the air-gap, dependent on angle, and map these regions of angles of incidence to regions of angles of departure according to Snell's law. Pale, blue dashed arrows show return paths for each beam. The shown beam trajectories are not the result of accurate computation but rather just for a qualitative representation.

refractive surface within the Aluminium reflector sheets. Because of the geometry of the set-up the entire fascicle energy will reach the photodetector after a certain number of bounces despite this multifurcation effect.

However, with every bounce between the pair of reflectors each 'child' beam born by the multifurcation process will give rise to its own 'grandchildren' beams via the same process of multifurcation caused by the presence of the Silicon dioxide layer. This implies that the more the beam bounces back and forth between the reflector layers the more the energy contained within it is spread between rapidly increasing numbers of significant component beams.

Using a technique similar to that used for analysing the transmittance mode case (Section 4.1.1.1) we can split these component beams into sets according to how many times they bounced within the Silicon dioxide layer before exiting into the Si region of the die as compared to the main component. This creates are few distinct classes of beam components: a) Type-a child beams that follow the path of the main beam with a few extra 'mini-bounces'[6] between the top reflector and the refractive surface between oxide and Silicon. b) Type-b child beams that instead of following the path of the main component suffer reflection at the refractive interface and return to Silicon, thereby suffering fewer mini-bounces than the main component. c) Type-c, mixed components that in some parts suffer extra mini-bounces whilst in others they suffer fewer than the main components. These types of components are illustrated in Figure 4.11.

For type-a components: We observe that the only difference between a beam that has bounced N times between the Aluminium reflectors and M times within the SiO_2 region before hitting the detector

[6]We will formally define a mini-bounce as any pair of ascending and descending trajectories between the $Si - SiO_2$ refractive surface and the top reflector regardless of whether they are a result of reflection or transmission through teh refractive surface.

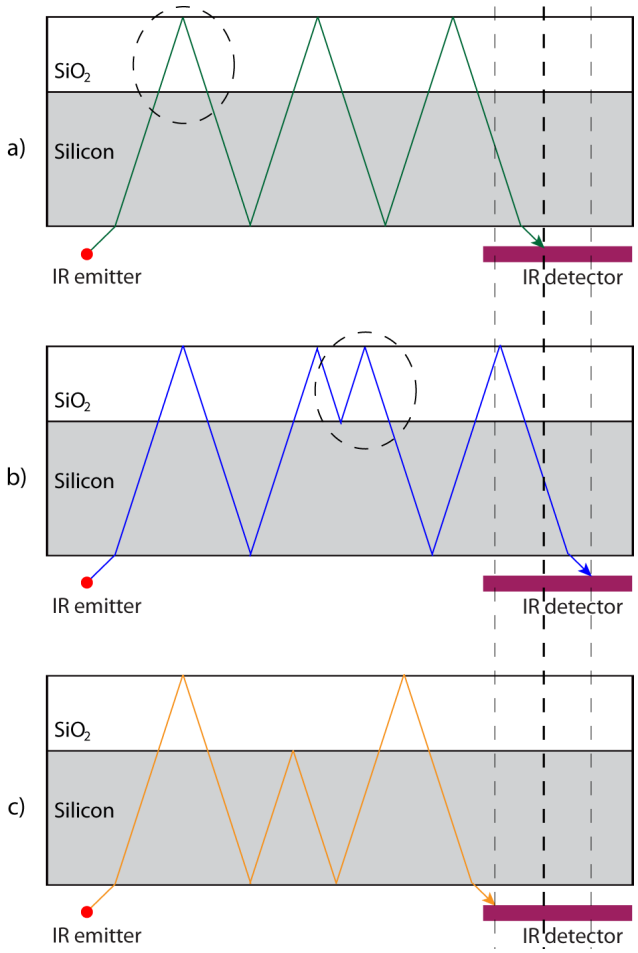

Figure 4.11: Three component beams illustrated against a simplified backdrop (Aluminium reflectors not shown): a) Main component. b) Type-a component with one extra mini-bounce (order 1). c) Type-b component with one missing mini-bounce (order 1). Dashed circles illustrate 'mini-bounces' stressing the fact that a pair of ascending and descending beam trajectories through the Silicon dioxide -in that order- constitutes a mini-bounce regardless of the specifics of how it has arisen. Thin, dashed vertical black lines shown the termination points of the beams compared to that of the main component (thick dashed line). Please note that where the extra or missing mini-bounces occur is irrelevant to the final lateral displacement but only the number of such additional or missing mini-bounces is significant. The internal structure of the Silicon region of the die and differences in the speed of light due to different refractive indices are not illustrated for clarity.

and another beam that has suffered N bounces through Si and M+1 within the SiO_2 is that the latter has suffered additional losses due to the extra encounter with the $Si - SiO_2$ surface. These (fractional) losses arise due to the fact that upon each encounter with the $Si - SiO_2$ surface part of the beam is transmitted further and part is reflected. These transmitted and reflected components represent the same fractions of the 'parent' beam (K_{mini}) for each pass at a given angle.

The above observation is important because it tells us that for any beam fascicle under study the main type-a components will be the ones that have suffered these additional 'mini-bounces' through the Silicon dioxide layer few times. Another important observation is that the number of 'mini-bounces' determines the lateral displacement between the 'main' component that hasn't suffered any mini-bounces and the component under study at the exit from the waveguide. Therefore regardless of the number of bounces through Silicon N, any beam that has been subjected to a single mini-bounce will exit the waveguide with exactly as much horizontal displacement with respect to the main component as that introduced by the mini-bounce; a small quantity as we've discussed in the reflectance mode section and the discussion on incidence angle region R_1.

The value of N though influences how many components have been subjected to exactly M mini-bounces. For the simple case of a single mini-bounce we have N such components, for two mini bounces we have $\frac{N(N-1)}{2}$ and generally for N bounces and M mini-bounces we have:

$$Q = \frac{N!}{M!(N-M)!} \tag{4.44}$$

where Q is the number of components subjected to M additional mini-bounces compared to the main component and the entire equation is a well-known combinatorial formula.

Thus, the values of Q and K_{mini} determine how much power is carried in various components of the fascicle. Specifically, the power carried in all components that have M extra mini-bounces compared to the main component is given by:

$$E_M = I_0 \frac{K_{mini}^M N!}{M!(N-M)!} \tag{4.45}$$

where E_M is the energy carried by the said components and notably $I_0 K_{mini}^M$ represents the energy of each individual component whilst the rest of the right hand side represents the number of suitable components.

Since this phenomenon is prevalent only in R_1 we can assume that horizontal displacement through the dioxide layer is small compared to that through Silicon and therefore generally all the beam energy within the fascicle will reach the photodetector with little lateral displacement compared to the 'main' component with respect to the scale of the horizontal displacement introduced by each bounce and will therefore reach the photodetector after N or N+1 bounces. To simplify the analysis we will ignore this 'edge effect[7]' and simply consider the case where all components reach the photodetector after N steps, in line with the main component (optimistic scenario). An illustration of the edge effect can be seen in Figure 4.12.

[7] Thus called because it concerns beam components that are distinguished by relatively little lateral displacement between them reaching the far edge of the bottom-side reflector at exactly the correct location for some of them to just about meet the reflector once more and others to graze by it and head straight towards the photodetector.

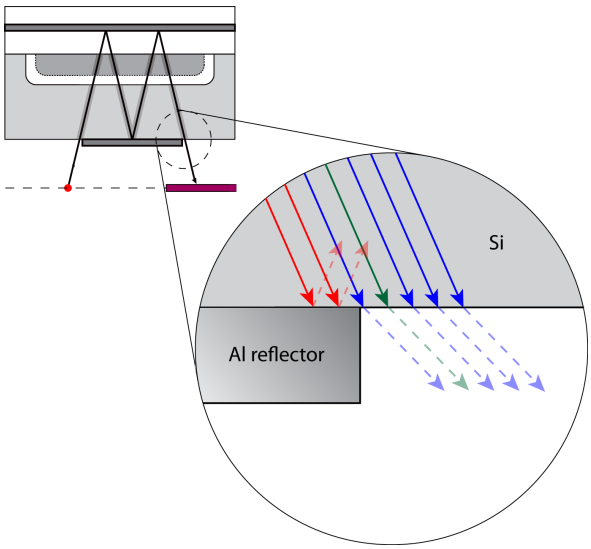

Figure 4.12: Edge effects can be difficult to handle analytically. The problem consists of the bundle of significant beam components reaching the end of the waveguide in very close proximity to the edge of the bottom reflector layer. Shown is the main component (green) along with some significant child component bundles (those within three mini-bounces of the main component). Of those, some hit the edge of the Aluminium waveguide and are subsequently reflected upwards towards the top surface of the die (red) whilst others are allowed to exit (or at least their transmitted part is - shown in blue). The top left side of the figure is a simplified version of 4.9.

Very similarly in scenario b): We have components that are subjected to fewer mini-bounces than the main component and will therefore reach the photodetector with smaller overall lateral displacement than the main component (ignoring edge effects). For these components it is difficult to define intensity as a function of mini-bounces that they miss because if we take any example component missing any number of mini-bounces and pair it with any component that consists of the example component plus one mini-bounce at some point A, then the example component will be arising from the reflected beam at A whilst its paired 'example plus' component will be arising from the transmitted beam at A.

Depending on angle the transmitted or the reflected component will be stronger upon starting the descending leg through Silicon. Specifically, the transmitted component will start the descending leg after being subjected to two transmission coefficient losses, one for each time it crosses the refractive surface and will therefore be described by:

$$I_T = I_0 T^2 \tag{4.46}$$

where I_T represents the transmitted component after it reenters Silicon and starts its descending leg and T represents the transmission coefficient that applies given the specifics of the geometry.

On the other hand the reflected component starts its descending leg after being subjected to the reflection coefficient:

$$I_R = I_0 R \tag{4.47}$$

where I_R is the reflected component and R the reflection coefficient given the specifics of the geometry.

If the following condition is satisfied:

$$T^2 > R \tag{4.48}$$

then the transmitted component carries more energy than the reflected component directly implying that the main component at all angles where the condition is satisfied is more significant than both components that have suffered more mini-bounces and fewer. The combinatorial formula 4.44 will apply to these lesser type-b components just as it does to their type-a counterparts, the only difference being that M will now represent a deficit of mini-bounces compared to the main component and not a surplus.

On the other hand, if the reverse applies, and $R < T^2$, then the most significant component is the one always reflected at the $Si - SiO_2$ interface (we will call this the 'short component') and the analysis done before for the main component and type-a components (those with extra mini-bounces) could be modified so that the short component plays the role of the main component and all other components are then type-a components with respect to it.

Dealing with mixed, type-C components is extremely difficult (unless in the case where criterion 4.48 is not met) and is better relegated to numerical solvers.

The important conclusion is that the total lateral displacement for all significant beam components, regardless of type will be fairly similar to that of the main component in the scale of the lateral displacement introduced by each main bounce and for angle region R_1 (as discussed for type-a components). This follows from the fact that in all cases the significant beam components will be given by the most significant component (be that the main or the minimum component) and components within relatively few extra (or minus) mini-bounces of the dominant component. This can be seen in figure 4.12 for example, where the main component and the bundles of components within three mini-bounces of the main component are shown only, the implicit assumption being that any components farther away than that carry an insignificant percentage of the energy that eventually reaches the photodetector.

Exceptions to this assumption might apply for certain angles (for example where condition 4.48 is just about met or just about not met) and definitive proof can only come if the losses as a function of extra or minus mini-bounces vs. the most significant component can be proven to increase faster than the number of components suffering that number of extra or minus bounces (at least beyond a certain, relatively small number of mini-bounces). The development of such detailed mathematical modeling is well outside the scope of this thesis.

The analysis in the rest of this section applies for the region of the parameter space where after a certain, relatively small number of mini bounces away from those suffered by the main component (whichever it is) the energy carried by those sub-components will become vanishingly small and that for angles of interest the overall lateral displacement of any set of beam components carrying any significant energy will be close to that of the main component. Therefore we can consider that all beams, regardless of type will reach the photodetector after the same number of bounces as the main component, if we are ignoring edge effects.

The net effect of the discussion so far is that we can discount the existence of the SiO_2 layer and proceed with analysing a simplified Air-Si model with two reflectors. We now have all the background necessary to start the beam-fascicle analysis proper.

The analysis begins by computing the maximum and minimum critical angles for this system, i.e. the critical angles outside whose range the beams either miss the waveguide (θ too small) or hit it but from the underside and are thus lost. Crucially, beams with angles above the maximum critical will be completely lost as they never even enter Silicon territory. On the other hand for angles below minimum critical the picture is more complicated. At 0^o, i.e. when the fascicle falls perpendicularly to the die surface, there is no lateral displacement with every bounce so the beam never enters the waveguide. For angles increasing between 0^o and the minimum critical value more and more components of the emitted fascicle enter the waveguide after fewer and fewer bounces between the top-side reflector and the un-plated, lossy bottom side of the die. Investigating the fine details of what fascicles (parameter: angle of departure) and what components of these fascicles (parameter: number of mini-bounces) enter the waveguide at what efficiency lies outside the scope of this thesis. As such we reach a key assumption of this section: *Beam fascicles emitted at less than the minimum critical angle will be ignored.*

To compute the maximum critical angle we simply take the arc-tangent of W_r/H_r:

$$\theta_M = \arctan\left(\frac{W_r}{H_r}\right) \tag{4.49}$$

where consistent with previous nomenclature θ_M is the maximum critical angle.

In order to find the minimum critical angle the trigonometry involved becomes more complicated. In order to simplify the analysis we once again ignore the additional displacement introduced by the presence of the SiO_2 layer as well as the additional components introduced by the 'mini-bounce' effect. We obtain:

$$W_r = H_r \tan(\theta) + 2D \tan(\theta_1) \tag{4.50}$$

where θ is the angle of departure and θ_1 is the angle of transmission through Silicon according to Snell's law and D is the thickness of the die. The first term of the sum is the displacement through the air-gap and the second term it the corresponding displacement through Silicon.

Using Snell's law equation 4.50 can be transformed into:

$$W_r = \frac{H_r \sin(\theta)}{\sqrt{1-\sin^2(\theta)}} + 2\frac{D\frac{n_s}{n_t}\sin(\theta)}{\sqrt{1-\left(\frac{n_s}{n_t}\right)^2 \sin^2(\theta)}} \tag{4.51}$$

where $n_s \approx 1$ (air) and $n_t \approx 3.5$ (Si) for the wavelengths of interest.

Solving equation 4.51 numerically with respect to θ will yield a value for the minimum critical angle θ_m.

For each fascicle emitted at a valid angle there will be a refracted angle which will determine how the beam will travel through Silicon, which we have already denoted as θ_1. The lateral displacement introduced in a beam by traversing the Silicon part of the die twice (once upwards, once downwards) will be given by:

$$\Delta W = 2D \tan(\theta_1) \tag{4.52}$$

At this point we need to define the horizontal distance between the point light source and the far edge of the bottom-side reflector. we will call this variable H_s. We observe a couple of things: First, the lateral displacement of any beam fascicle from the point light source to the point where it exits the die from the underside will always be greater than or equal to H_s. Second: the lateral displacement of a beam from emitter to the aforementioned exit point will be a discrete set of points given by:

$$\Lambda(k) = H_r \tan(\theta) + 2kD \tan(\theta_1) \geq H_s \tag{4.53}$$

where $\Lambda(k)$ is the emitter to exit point lateral displacement of the beam and the inequality part of the expression marks the first observation in the pair. The first term of the sum represents the lateral displacement through the air-gap while the second term represents the total lateral displacement through Silicon.

For each fascicle there will be exactly one minimum $\Lambda(k)$ that satisfies the inequality from 4.53 and it will be a function of angle of departure. For each such value there will be a corresponding k value, which is important because the factor $2k$ indicates the number of passes through the Silicon portion of the die achieved by the beam en route from the photoemitter to the photodetector.

The rest of the analysis is lifted straight from the reflectance configuration: A coefficient of fractional loss per pass as a function of bias voltage $K_{Si}(V_{bias})$ is defined and then the computed remaining intensity (after losses) for all fascicles within the valid angle interval $[\theta_m, \theta_M]$ are integrated over angle in order to obtain the final light intensity reaching the photodetector. Specifically for the received beam intensity of a fascicle we obtain the relation:

$$\boxed{I_{det} = \gamma I_0 \,|\, \gamma = \left(\frac{T_m + T_e}{2}\right)_{asc} K_{Si}^{2k}} \tag{4.54}$$

where the equation follows the form of equation 4.30. The first factor of the product refers to radiation transmitted upon its first encounter with the Air-Si interface. Any radiation reflected at this point is assumed lost. The second factor is the lumped loss coefficient raised to the power of $2k$ in order to represent the $2k$ passes of the beam through the Silicon. Finally, because of our assumption of infinite length top-side reflector and detector, we can drop the last factor of equation 4.30. The radiation reflected upwards again instead of carrying on through the lossy Si-Air interface towards the photodetector will be subjected to more bouncing, extra modulations and will eventually all reach the photodetector with a few extra K_{Si} factors in its modulation expression. These are discounted for simplicity (conservative calculation).

Equation 4.54 can also be expressed in difference form with respect to bias voltage:

$$\boxed{\Delta I_{det} = I_0 \left(\frac{T_m + T_e}{2}\right)_{asc} (K_{Si}(V_0)^{2k} - K_{Si}(V_1)^{2k})} \tag{4.55}$$

where V_0 and V_1 are different bias voltages.

Finally, for a non-uniform point light source in 2D we can define a function that describes emitted light intensity as a function of angle of departure $Z(\theta)$. In order to find the amount of light power from

all fascicles reaching the photodetector as a function of bias voltage we simply apply:

$$P_{2D} = \int_{\theta_m}^{\theta_M} [Z(\theta)\gamma(\theta)]^2 d\theta \tag{4.56}$$

which of course is a function of bias voltage. The underlying analysis was performed in more detail in the reflectance configuration section.

Finally, generalising the bounce mode set-up to 3D follows exactly the same pattern as the corresponding generalisation for the reflectance mode case if the 3D geometry is assumed to be generated by a 360^o rotation of the 2D system about the 0^o axis that intercepts the point light source. As such, this generalisation will not be explicitly discussed in this section.

4.2 Worked example: transmittance mode

We shall now show how a simple, approximate numerical solution as to the degree of absolute modulation depth can be provided for a simple example of a transmittance configuration system. Note: This is a highly conceptual example using 'well-behaved' abrupt junctions where the aim is to show the mathematical mechanics of the system and perform some useful analysis on the resulting equations. It is not intended as a method of predicting or estimating performance in real, CMOS pn-junctions.

We shall begin by defining the parameters of the set and the proceed to calculate the resulting solution. Next, we shall use the resulting equation to provide some insight as to how the results depend on various key parameters. Finally, we shall give some rough results based on an even more simplified system set-up in order to illustrate dependence of modulator performance on doping concentrations.

4.2.1 System set-up.

Let us assume we have a transmittance-mode set-up of active modulation area[8] $A = 1\,\mu m^2$, die thickness $D = 300\,\mu m$, diffusion layer thickness $d = 2\,\mu m$, Silicon dioxide layer thickness in the few micron range, and that the doping concentration in both substrate and diffusion layer is completely uniform throughout their entire illuminated areas (i.e. the junction formed between them is abrupt) with doping concentrations of $10^{17}\,dopants/cm^3$ (substrate) and $10^{20}\,dopants/cm^3$ (diffusion layer). Additionally, we know that the wavelength is $1550\,nm$. Finally, we switch the modulator between bias voltage of $0\,V$ and $5\,V$ reverse bias.

This information allows us to calculate the free-carrier absorption coefficient in the substrate as $3 \times 10^{-5}\,\mu m^{-1}$ (coefficient for free holes at $10^{17}\,dopants/cm^3$ at $\lambda = 1550\,nm$) and in the diffusion layer as $10^{-1}\,\mu m^{-1}$ (coefficient for free electrons at $10^{20}\,dopants/cm^3$ at $\lambda = 1550\,nm$). Moreover, we know that in the wavelength band of interest SiO_2 shows an absorption coefficient of approx. 10^{-7} - $10^{-6}\,cm^{-1}$ ([1]), which renders SiO_2 effectively transparent for layers of a few microns thick as encountered in CMOS.

[8] The active modulation area is computed as the intersection of the light source area with the modulator area and the photodetector area if we collapse all their areas on the same plane in the direction of light travel. In simple term, the intersection of emitter, modulator and detector area 'as seen from above' since the light is travelling perpendicularly to these structures by assumption.

Finally we set the temperature such that the intrinsic carrier concentration for Si conveniently becomes equal to $10^{10}\,dopants/cm^3$. Thus absolute temeprature T is equal to $298\,K$.

The entire set of relevant parameters is summarised in table 4.2.

Table 4.2: Parameters used for the worked example in this section. A: active modulation area. D: Die front-end thickness. d: Diffusion layer thickness. N_A: Substrate doping concentration ('acceptor' concentration). N_D: Diffusion layer doping concentration ('donor' concentration). n_i: Intrinsic carrier concentration. λ: Radiation wavelength. a_{sub}: Substrate absorption coefficient. a_{diff}: Diffusion layer absorption coefficient. a_{SiO2}: Silicon dioxide absorption coefficient. V_1: Bias voltage 1. V_2: Bias voltage 2. I_0: Irradiance at departure from light source. T: Absolute temperature.

PARAMETER	VALUE	UNITS
A	1	μm^2
D	300	μm
d	2	μm
N_A	10^{17}	$dopants/cm^3$
N_D	10^{20}	$dopants/cm^3$
n_i	10^{10}	$dopants/cm^3$
λ	1550	nm
a_{sub}	4×10^{-5}	μm^{-1}
a_{diff}	2×10^{-1}	μm^{-1}
a_{SiO2}	approx. 0	μm^{-1}
V_1	0	V
V_2	-5	V
I_0	1	$\mu W/\mu m^2$
T	298	K

4.2.2 Solution

We begin the analysis by computing the losses through the substrate and the losses through the diffusion layer. Using basic semiconductor theory we find that the depletion region width W is given by (see [7] and chapter 3):

$$W = \sqrt{\frac{2\epsilon}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)\left(\frac{kT}{q} ln\left(\frac{N_A N_D}{n_i^2}\right) - V_{bias}\right)} \tag{4.57}$$

where V_{bias} is the external, forward-bias voltage and all the rest of the parameters are either universal constants (ϵ, q, k) or have already been specified in our parameter table (4.2).

For zero-bias, this value is computed to be equal to $114.6\,nm$. we can now use the charge neutrality equation for abrupt junctions (see equation 3.50 on 76 for background behind this equation):

$$W_p N_A = W_n N_D \tag{4.58}$$

where W_p and W_n are the depletion region widths in the substrate (p-type) and diffusion layer (n-type) respectively.

We combine equation 4.58 with the fact that the total depletion region width equals the substrate-side component plus the diffusion layer-side component:

$$W_p + W_n = W \tag{4.59}$$

to obtain:

$$Wp = \frac{\frac{N_D}{N_A} W}{1 + \frac{N_D}{N_A}} \tag{4.60}$$

We can now calculate a zero-bias depletion region width in substrate equal to $114.5\,nm$ and corresponding depletion width in the diffusion region of approx. $0.1145\,nm$. For similar calculations performed for $5\,V$ reverse bias the numbers become $278.9\,nm$ for the substrate-side and approx. $0.2789\,nm$ for the diffusion region side. We summarise these in table 4.3.

Table 4.3: Calculated depletion region widths for the two chosen bias voltages.

PARAMETER	**VALUE**	**UNITS**
$W_p(0)$	114.5	nm
$W_n(0)$	0.1145	nm
$W_p(-5)$	278.9	nm
$W_n(-5)$	0.2789	nm

where $W_x(y)$ stands for depletion region width in region x (p: substrate or n: diffusion) under forward bias of y Volts.

The next step is to simply calculate electro-magnetic beam path length through the substrate and through the diffusion layer which is trivially done by subtracting the depletion region widths from their host layers' overall thicknesses. We obtain:

$$L_0 = D - d - W_p(V_{bias}) = 297.8855\,\mu m|(V_{bias} = 0\,V) = 297.7211\,\mu m|(V_{bias} = -5\,V) \tag{4.61}$$

for the substrate-side path length and:

$$L_{N0} = d + W_n((V_{bias}) = 1.9998855\,\mu m|(V_{bias} = 0\,V) = 1.9997211\,\mu m|(V_{bias} = -5\,V) \tag{4.62}$$

for the diffusion layer path length.

We can now compute the transmittance of a beam of arbitrary intensity through both substrate and diffusion layer. Since both substrate and diffusion layer are uniformly doped we will end up with Beer-Lambert form expressions:

$$K_{sub} = e^{-a_{sub}L_0} = 98.8155\,\%|(V_{bias} = 0\,V) = 98.8162\,\%|(V_{bias} = -5\,V) \tag{4.63}$$

for the substrate and:

$$K_0 = e^{-a_{diff}L_{N0}} = 67.0336\,\%|(V_{bias} = 0\,V) = 67.0358\,\%|(V_{bias} = -5\,V) \tag{4.64}$$

for the diffusion layer.

Notably, the diffusion layer is the main cause of optical losses despite its relatively small thickness. We can now aggregate K_0 and K_{sub} in order to obtain the overall transmittance coefficient per pass through the die K_{Si}:

$$K_{Si} = 66.2396\,\%|(V_{bias} = 0\,V) = 66.2422\,\%|(V_{bias} = -5\,V) \tag{4.65}$$

We can now return to equation 4.13 on page 114 and use our computed values for K_{Si}, our pre-computed values for the various refractive index transmittance coefficients (see table 4.1 on page 111) in order to compute how how much of the initial intensity reaches the photodetector through the main component (transmitted through all refractive interfaces):

$$Q_0 = T_1 K_{Si} T_2 T_3 = 36.8568\,\%|(V_{bias} = 0\,V) = 36.8582\,\%|(V_{bias} = -5\,V) \tag{4.66}$$

where T_1, T_2 and T_3 are the transmittance coefficients at the $Air - Si$, $Si - SiO_2$ and $SiO_2 - Air$ interfaces respectively.

For first order components (a single mini-bounce) we can use equations 4.14 to 4.16 on page 114 and onwards to obtain similarly:

$$Q_A = 0.8021\,\%|(V_{bias} = 0\,V) = 0.8022\,\%|(V_{bias} = -5\,V) \tag{4.67}$$

$$Q_B = 0.2359\,\%|(V_{bias} = 0\,V) = 0.2359\,\%|(V_{bias} = -5\,V) \tag{4.68}$$

$$Q_C = 0.1415\,\%|(V_{bias} = 0\,V) = 0.1415\,\%|(V_{bias} = -5\,V) \tag{4.69}$$

Summing all these contributions together we see that the main and first order components of the beam allow for the passage of $Q_{0-1}(0) = 38.0363\,\%$ of the initial light intensity at zero-bias and $Q_{0-1}(-5) = 38.0378\,\%$ of the initial intensity at $5\,V$ reverse bias. $Q_{a-b}(x)$ stands for transmittance of 'ath' to 'bth' order components under x volts forward bias. This forms a lower bound for the amount of energy reaching the photodetector.

An upper bound can be found by performing similar calculations for components that do not reach the photodetector. We shall only consider the main components here, namely the ones that exit the die travelling in a downward direction (towards the light source) after having suffered exactly one reflection at one of the refractive interfaces:

$$P_A = R_1 = 31.0\,\% \tag{4.70}$$

for the component being reflected at the $Air - Si$ interface (simply the reflection coefficient at that interface).

$$P_B = T_1 K_{Si} R_2 K_{Si} T_1 = 3.3424\,\%|(V_{bias} = 0\,V) = 3.3426\,\%|(V_{bias} = -5\,V) \tag{4.71}$$

for the component reflected at the $Si - SiO_2$ interface.

$$P_C = T_1 K_{Si} T_2 R_3 T_2 K_{Si} T_1 = 0.5896\,\%|(V_{bias} = 0\,V) = 0.5896\,\%|(V_{bias} = -5\,V) \tag{4.72}$$

for the component reflected at the $SiO_2 - Air$ interface.

Together the first order components account for losses tantamount to 34.9320 % of the initial beam intensity at zero-bias and 34.9323 % under $5\,V$ reverse bias. Thus the upper bound as given by this calculation is approximately 65 % overall transmittance rate. Note, however, that this upper bound calculation does not take into account light lost through the Silicon and is therefore an overestimate. A new set of formulae dealing with losses would need to be developed in order for this estimate to become more accurate.

Continuing from the transmittance of the zeroth (main) and first order components we now need to multiply the $Q_{0-1}(0)$ and $Q_{0-1}(-5)$ values with the irradiance of the light source and the active modulation area in order to find how much power is being transmitted to the photodetector by the aforementioned zeroth and first order components under both bias conditions. We thus compute that the transmitted power is:

$$Z(V_{bias}) = I_0 A Q_{0-1}(V_{bias}) = 380.363\,nW|(V_{bias} = 0\,V) = 380.378\,nW|(V_{bias} = -5\,V) \tag{4.73}$$

This yields an absolute difference of approx. $15\,pW$, or approx. $39\,ppm$ of the intensity reaching the photodetector. We remind the reader that this is a lower bound yielded by a very simplified analysis.

4.2.3 Dependence of absolute modulation depth on key parameters

From the previous sub-section we notice that the absolute modulation depth lower boundary can be expressed as:

$$\Delta Z(V_1, V_2) = |Z(V_1) - Z(V_2)| = I_0 A|Q_{0-1}(V_1) - Q_{0-1}(V_2)| \tag{4.74}$$

This can then be further 'unfolded' to yield:

$$\Delta Z(V_1, V_2) = I_0 A|Q_0(V_1) + Q_A(V_1) + Q_B(V_1) + Q_C(V_2) - Q_0(V_2) + Q_A(V_2) + Q_B(V_2) + Q_C(V_2)| \tag{4.75}$$

which, using equations 4.13 - 4.16 further unfolds into:

$$\boxed{\Delta Z(V_1, V_2) = I_0 A|o_1(K_{Si}(V_1) - K_{Si}(V_2)) + o_3(K_{Si}^3(V_1) - K_{Si}^3(V_2))|} \tag{4.76}$$

where $o_1 = (T_1T_2T_3)(1 + R_2R_3)$, $o_2 = T_1R_1T_2T_3(R_2 + R_3T_2^2$ are constants determined solely by the refractive indices of the media involved in the optical path.

We also know that K_{Si} is given by:

$$\boxed{K_{Si}(V) = K_{sub}(V)K_0(V) = e^{-(a_{sub}L_0(V)+a_{diff}L_{N0}(V))}} \tag{4.77}$$

$L_0(V)$ is given by:

$$\boxed{L_0 = D - d - W_pV_{bias} = D - d - \frac{\frac{N_D}{N_A}\sqrt{\frac{2\epsilon}{q}\left(\frac{N_A+N_D}{N_AN_D}\right)\left(\frac{kT}{q}ln\left(\frac{N_AN_D}{n_i^2}\right) - V_{bias}\right)}}{1 + \frac{N_D}{N_A}}} \tag{4.78}$$

from equations 4.57 on page 136 and 4.60 on page 137.

Finally L_{N0} can be computed as:

$$\boxed{L_{N0} = d - W_nV_{bias} = d - \frac{\frac{N_A}{N_D}\sqrt{\frac{2\epsilon}{q}\left(\frac{N_A+N_D}{N_AN_D}\right)\left(\frac{kT}{q}ln\left(\frac{N_AN_D}{n_i^2}\right) - V_{bias}\right)}}{1 + \frac{N_A}{N_D}}} \tag{4.79}$$

by applying the same methods we used in order to determine L_0.

By considering equations 4.76-4.79 we can begin to discern the influence of various key parameters on system performance, namely: a) voltage bias, b) die and diffusion layer thicknesses and c) the irradiance of the light source.

Bias voltage dependence

The voltage bias feeds into equations 4.78 and 4.79 where it changes the values of the optical path lengths in accordance to the square root of the built-in potential (the $V_{bi} = \frac{kT}{q}\left(\frac{N_A+N_D}{N_AN_D}\right)$ factor) plus the reverse bias ($-V_{bias}$). Thus:

$$\frac{\partial L_0}{\partial V_{bias}} \propto \frac{1}{\sqrt{V_{bi} - V_{bias}}} \tag{4.80}$$

The same expression is also found for L_{N0}.

Since the path lengths feed into equation 4.77 we can easily determine the relation between voltage bias and transmittance through each pass through the die:

$$\boxed{K_{Si} \propto e^{C_0\sqrt{V_{bi}-V_{bias}}}} \tag{4.81}$$

where C_0 is a constant depending on doping concentrations, device geometry and absorption coefficients (themselves a function of doping concentrations mobility etc. see equation 4.1 on page 107). Additionally by differentiating K_{Si} by V_{bias} we can determine that:

$$\frac{\partial K_{Si}}{\partial V_{bias}} \propto \frac{e^{C_0\sqrt{V_{bi}-V_{bias}}}}{\sqrt{V_{bi}-V_{bias}}} \tag{4.82}$$

$$\frac{\partial K_{Si}^2}{\partial V_{bias}} = 2K_{Si}\frac{\partial K_{Si}}{V_{bias}} \propto \frac{e^{2C_0\sqrt{V_{bi}-V_{bias}}}}{\sqrt{V_{bi}-V_{bias}}} \tag{4.83}$$

$$\frac{\partial K_{Si}^3}{\partial V_{bias}} = 3K_{Si}^2\frac{\partial K_{Si}}{V_{bias}} \propto \frac{e^{3C_0\sqrt{V_{bi}-V_{bias}}}}{\sqrt{V_{bi}-V_{bias}}} \tag{4.84}$$

Finally, we can see that for small differences (enough for linearisation of K_{Si}, K_{Si}^2 and K_{Si}^3) the absolute modulation depth $\Delta Z(V_1, V_2)$ will depend on the choice of operating bias voltage thus:

$$\boxed{\Delta Z(V_1, V_2) \propto C_1\left(\Delta V\frac{e^{C_0\sqrt{V_{bi}-V_{bias}}}}{\sqrt{V_{bi}-V_{bias}}}\right) + C_2\left(\Delta V\frac{e^{3C_0\sqrt{V_{bi}-V_{bias}}}}{\sqrt{V_{bi}-V_{bias}}}\right)} \tag{4.85}$$

where C_1 and C_2 are constants that depend on doping concentrations, device geometry, absorption coefficients, the light source irradiance and the refractive indices of the materials involved.

Material thicknesses dependence

The dependence of electro-optical modulation on material thicknesses (die thickness and diffusion layer thickness) is far more straightforward. WE begin by noticing that the path length dependences take the form:

$$\frac{\partial L_0}{\partial D} = 1 \tag{4.86}$$

$$\frac{\partial L_0}{\partial d} = -1 \tag{4.87}$$

$$\frac{\partial L_{N0}}{\partial d} = 1 \tag{4.88}$$

Let us begin by examining the die thickness for constant diffusion layer width first (D). $\frac{\partial L_0}{\partial D}$ feeds into equation 4.77 thus resulting in:

$$\boxed{K_{Si} \propto e^{-a_{sub}D}} \tag{4.89}$$

which links transmittance for each pass through the die to D. Furthermore for the partial derivative we obtain:

$$\frac{\partial K_{Si}}{\partial D} \propto e^{-a_{sub}D} \tag{4.90}$$

Applying the techniques used for the derivation of equation 4.85 (linearisation around operating point) we derive that the absolute modulation depth depends on D in accordance to:

$$\boxed{\Delta Z(V_1, V_2) \propto C_3 \left(\Delta V e^{-a_{sub}D}\right) + C_4 \left(e^{-3a_{sub}D}\right)} \tag{4.91}$$

with C_3 and C_4 constants.

On the other hand, if we consider modulation as a function of the diffusion layer width d we obtain a K_{Si} of the form:

$$\boxed{K_{Si} \propto e^{(a_{sub}-a_{diff})d}} \tag{4.92}$$

which links transmittance for each pass through the die to d. At the same time for the partial derivative of K_{Si} we have:

$$\frac{\partial K_{Si}}{\partial d} \propto e^{(a_{sub}-a_{diff})d} \tag{4.93}$$

Similarly to the derivation of 4.85 we can now express the dependence of absolute modulation depth on d thus:

$$\boxed{\Delta Z(V_1, V_2) \propto C_5 \left(\Delta V e^{(a_{sub}-a_{diff})d}\right) + C_6 \left(e^{3(a_{sub}-a_{diff})d}\right)} \tag{4.94}$$

with C_5 and C_6 constants.

Light source irradiance dependence

This is the most straightforward case since I_0 only appears in equation 4.76 and affects ΔZ in direct proportionality. Thus we can state:

$$\Delta Z \propto I_0 \tag{4.95}$$

4.2.4 Simplified case: symmetrical, abrupt and homogeneous junction

In this section we shall display simple results obtained from a very simplified system whereby we have a symmetric, abrupt and homogeneous pn-junction. The thickness of both p- and n-type sides is the same

as is their doping concentration. The objective is to illustrate how fractional modulation depth changes in response to changes in the common doping concentration that characterises both sides of the device. We use absorption coefficient values for $\lambda = 1550\,nm$ and assume the intrinsic carrier concentration of Silicon to be $n_i = 10^{10}\,carriers/cm^3$ at the temperature of interest (around $278\,K$).

For this section we shall be using the approximation that fractional modulation depth $\frac{\Delta T}{T}$ is given by:

$$\frac{\Delta T}{T} \approx aw \tag{4.96}$$

where a is the average absorption coefficient between electrons and holes at some given, common doping concentration and w is the difference between the depletion region widths at the 'ON' and 'OFF' bias points. This approximation is derived in the theoretical background chapter (section 3.2.3).

For six different common doping concentration levels we obtain the results in table 4.4. The absorption coefficient values are approximations of measurements by Soref et al. [3] whilst depletion region widths at various bias voltages, the built-in potential and fractional modulation depth are all computed values. The dependence of fractional modulation depth on doping concentration can be seen in Figure 4.13.

Table 4.4: Electro-optical modulation depth and key parameters associated with modulation performance for a few simple cases of symmetrical (same thickness, same doping concentration), abrupt, homogeneous pn-junctions. Legend: N: doping concentration levels, common to both p- and n-type sides of the junction. V_{bi}: Built-in potential. $W(x)$: depletion region width under x volts forward bias. ΔW: $W(-5) - W(0)$. a_x: Absorption coefficient for holes ($x = h+$), electrons ($x = e-$) and the average between the two ($x = avg$). $\frac{\Delta T}{T}$: Modulation depth. Dp/cm^3: Dopants per cubic cm. Cr/cm^3: Free carriers per cubic cm.

N	V_{bi}	W(0)	W(-5)	ΔW	a_{h+}	a_{e-}	a_{avg}	$\frac{\Delta T}{T}$
Dp/cm^3	V	nm	nm	nm	μm^{-1}	μm^{-1}	μm^{-1}	ppm
10^{20}	1.197	5.6	12.7	7.1	10^{-4}	2×10^{-4}	1.50×10^{-4}	1065.00
10^{19}	1.078	16.7	39.7	23.0	7.00×10^{-6}	1.3×10^{-5}	10^{-5}	230.00
10^{18}	0.958	49.7	124.1	74.4	4.90×10^{-7}	8.45×10^{-7}	6.68×10^{-7}	49.66
10^{17}	0.838	147.0	388.6	241.6	3.43×10^{-8}	5.49×10^{-8}	4.46×10^{-8}	10.78
10^{16}	0.718	430.4	1216.4	786.0	2.40×10^{-9}	3.57×10^{-9}	2.99×10^{-8}	2.35
10^{15}	0.599	1242.5	3806.2	2563.7	1.68×10^{-10}	2.32×10^{-10}	2.00×10^{-10}	0.51

Notably, modulation depth improves with higher doping concentrations and the slope of the log-log plot is approximately 0.66, which is slightly higher than the value of 0.5 expected if the fractional modulation depth depended on the square root of the doping concentration. The reason for this is the fact that our simplified model ignores built-in potential dependence on doing concentration; a dependence that alters how depletion region width difference w changes with doping concentration. For details see the relevant theoretical background chapter section 3.2.3.

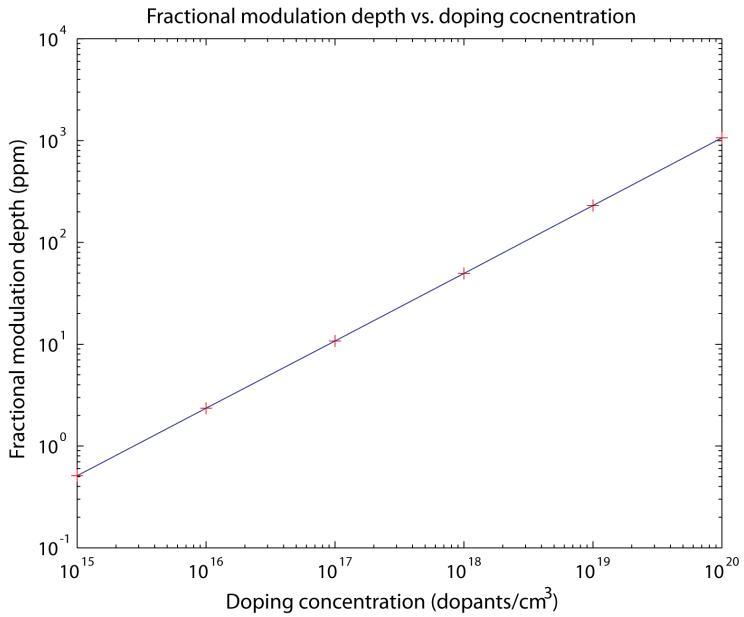

Figure 4.13: Dependence of fractional modulation depth on doping concentration for a simple test set-up featuring a symmetrical (same thickness for p- and n-type sides and symmetrical doping concentration profiles), abrupt and homogeneous pn-junction. Higher doping concentrations lead to improved modulation depths.

4.3 Worked example: semi-realistic CMOS pn-junction in transmittance mode

In this section we shall apply the techniques used previously in 4.2 in order to try and predict how a realistic CMOS pn-junction will modulate light passing through it. In contrast to the previous example in section 4.2, this is not intended as a largely mathematical exercise, but an attempt to (albeit very roughly) estimate modulation performance in a realistic pn-junction with as much detail included as is practical within the time constraints of the project.

For the purposes of this modelling activity we shall begin by setting up a CMOS pn-junction and a biasing strategy, then proceed to analyse its structure and determine the nature and extent of the depletion region. Armed with this information we can then proceed to compute how the areal and side-wall components of the pn-junction modulate light and predict the performance of the device, both as a whole and by component contributions.

4.3.1 Set-up

We begin by setting up the problem. Suppose we have a very simple, but realistic N-well on substrate pn-junction. Geometrically it will consist of three regions: a) the areal junction, b) the side-wall and c) the transitional region between the two. The extent of the device will be 10×10 microns (nominal, or 'mask' extent; in reality the junction will be wider due to straggle). The die is 300 microns thick excluding back-end.

pn-junction set-up:

The device will be formed by four ideal ion implantation processing steps in accordance to equation 3.97 found on page 92 in section 3.3.2, extended to account for the 3D structure of the device. The substrate will be considered as uniformly doped at 10^{16} $dopants/cm^3$Thus the n-type doping profile $N_Z(x, y, z)$ (doping concentration as a function of each point in the volume of the device) will be given by:

$$N_Z(x,y,z) = \sum_i \frac{A_0}{2} e^{\left(-\frac{(x-x_\mu)^2}{\sigma_x^2}\right)} \left[erfc\left(\frac{y-\zeta}{\sqrt{2}\sigma_y}\right) - erfc\left(\frac{y+\zeta}{\sqrt{2}\sigma_y}\right)\right]\left[erfc\left(\frac{z-\zeta}{\sqrt{2}\sigma_z}\right) - erfc\left(\frac{z+\zeta}{\sqrt{2}\sigma_z}\right)\right] \tag{4.97}$$

where x denotes depth and y, z denote directions parallel to the surface of the die. Index i shows the independent doping processing steps involved in creating the junction. Lateral doping concentration spread parameters σ_y and σ_z (mutually orthogonal) will be equal and together with the value of longitudinal doping concentration spread parameter σ_x common to all doping processing steps. The peak doping concentration depths for each doping step (parameters $x_{\mu,i}$) will be uniformly distributed between the surface ($x = 0$) and a depth of $1500\,nm$. Finally, peak doping concentration for each doping step A_0 will be common to all doping steps. All parameter values used for modelling our pn-junction doping profile are summarised in table 4.5.

Table 4.5: Parameters used for setting up our idealised CMOS pn-junction. Param.: parameter symbol.

Param.	**Value(s)**	**Unit**	**Description**
A_0	0.75×10^{19}	$dopants/cm^3$	Doping step peak doping concentration
i	{1, 2, 3, 4}	-	Index of ideal doping steps
$x_{\mu,i}$	$\{0, 50, 100, 150\}$	nm	Peak doping concentration depth for each doping processing step
σ_x	400	nm	Longitudinal doping concentration spread
σ_y, σ_z	100	nm	Lateral straggle doping concentration spread parameter
ζ	5000	nm	Half nominal device width in both lateral directions

Given equation 4.97 and the information contained in table 4.5 we can now numerically compute the doping profile of the pn-junction under consideration. The results of MATLAB simulation are shown in Figure 4.14.

Biasing and illumination regimes

For this example we shall use an 'OFF' bias voltage of $0\,V$, an 'ON' bias voltage of $-5\,V$ ($5\,V$ reverse bias).

We shall illuminate the entire device plus a $1\,\mu m$ thick 'frame' around it, so the illumination is an 11×11 micron square that shares its centre with the centre of the test device. The illumination will be a uniform, $1\,\mu W/\mu m^2$, monochromatic beam at $1550\,nm$ wavelength.

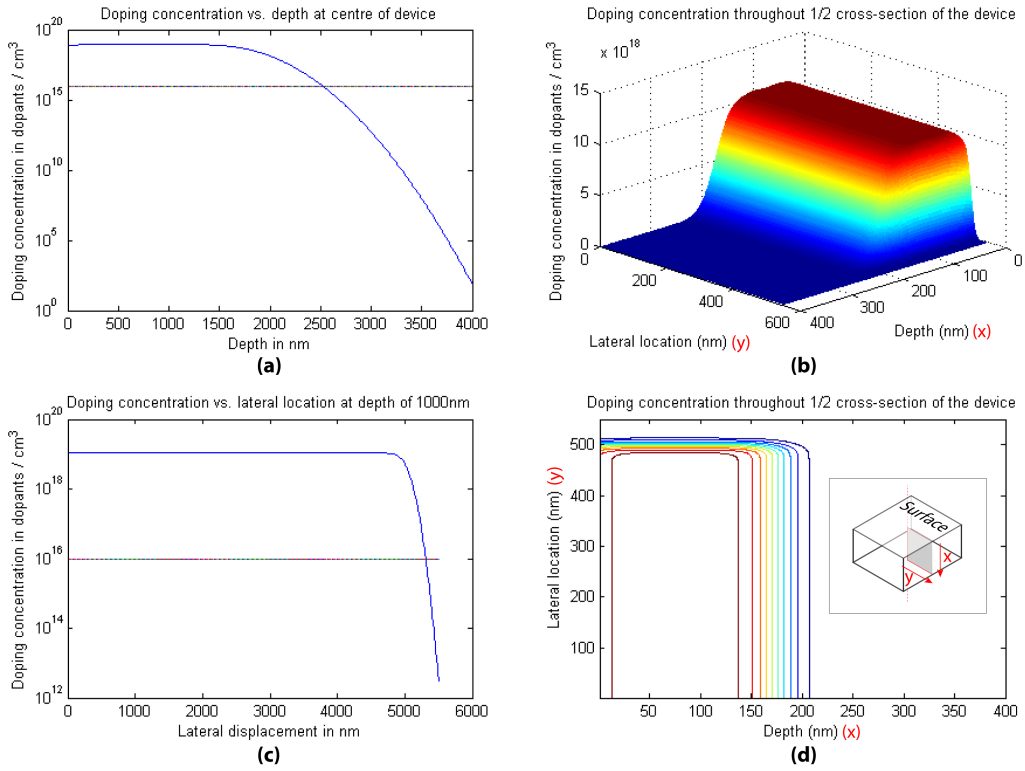

Figure 4.14: Plots exposing the semiconductor properties of the model pn-junction used for the 'modulation in CMOS' worked example. a) Doping concentration vs. depth at the centre of the device plotted on a logarithmic scale. The blue line shows the n-type dopant concentration that forms the well whilst the dotted line shows the baseline p-type dopant concentration present throughout the substrate. b) Linear surface plot of n-type concentration throughout a half-cross section of the device extending from the centre of the device (lateral location 0) perpendicularly to the nearest side-wall. The plane over which doping concentrations are plotted is shown in the inset of (d), gray shaded area. c) Same as (a), but the concentration profile is taken along a line extending from the centre of the device at a depth of $100\,nm$ to the nearest side-wall. d) Contour plot corresponding to (b). We notice how the areal and side-wall junction components are seprated fairly well.

4.3.2 Depletion region analysis

The next step in the model analysis is to study the nature of the depletion regions being formed in various subsets of the metallurgical surface. At this point we make a simplification: all junction components will be considered to be linearly graded in the immediate surroundings of the metallurgical surface so that we can use linearly graded models for the depletion region (not included in this thesis, but easily found in [7]). Specifically, we shall assume that in the vicinity of the depletion region the p-type doping concentration stays constant whilst the n-type dopant concentration gradient is constant and perpendicular to the metallurgical surface.

Areal junction

From the doping profile simulator we find that the metallurgical surface is situated at a depth of $2530\,nm$ and has a gradient at the centre of the device equal to approx. $\xi_{area} = 1.14 \times 10^{14}\ dopants/cm^3/nm$[9].

We now turn to Sze [7] and use the formulae therein for the computation of the built-in potential and the depletion region width for a given doping concentration gradient to find zero-bias values of V_{bi} and W:

$$V_{bi} = \frac{2}{3} V_T ln\left(\frac{\xi_{area}^2 \epsilon_{Si} V_T}{8 q n_i^3}\right) - 0.075 \approx 0.578\,V \tag{4.98}$$

where V_{bi} is the built-in potential across the linearly graded junction, V_T the thermal voltage at $T = 278\,K^o$, ϵ_{Si} the absolute electrical permittivity of Silicon q the fundamental unit of charge and n_i the intrinsic carrier concentration of Silicon at $T = 278\,K^o$. Note the final term in the right hand side of equation 4.98; it represents a correction factor introduced in [7] that better matches numerical results to experimental data. No explanation for the discrepancy is offered in the original source, however.

Next we compute the depletion region width:

$$W = \frac{2 n_i}{\xi_{area}} e^{\frac{V_{bi}}{2V_T}} \approx 12.4\,nm \tag{4.99}$$

where W is the depletion region width (both sides).

Given the computed net doping concentration gradient for this junction we expect the net doping at both edges of the zero-bias depletion region of this pn-junction to be approximately equal to $N_0 \approx 0.7 \times 10^{15}\ dopants/cm^3$.

Next, we need to perform an analysis of the depletion region under bias. [7] states that the depletion region width is proportional to the quantity $(V_{bi} - V_{bias})^{1/3}$. Thus, the ratio of the depletion region width under reverse bias of $5\,V$ over the same metric under zero-bias, R, will be given by:

$$R = \frac{W_{-5}}{W_0} = \left(\frac{V_{bi} + 5}{V_{bi}}\right)^{1/3} \approx 2.129 \tag{4.100}$$

[9] Value obtained by direct extraction from simulated data shown in Figure 4.14-(a). Equal to the computed gradient between the two neighbouring points that straddle the doping concentration of the substrate, $10^{16}\ dopants/cm^3$

where W_x is the depletion region width under a forward bias of x Volts. We notice how R depends on semiconductor properties through V_{bi}.

From equation 4.100 we obtain a depletion region width of approx. $26.4\,nm$ under our chosen reverse bias. At that distance from the metallurgical surface, the net doping concentration is equal to approx. 1.5×10^{15} $dopants/cm^3$. Thus, the average doping concentration throughout the depletion region edge swept volume[10] will be equal to approx. 1.1×10^{15} $dopants/cm^3$. This applies to both sides of the junction as it is perfectly symmetrical under our linear grading assumptions.

For simplicity, we shall assume that the entire areal junction is adequately characterised by these numbers, i.e. is perfectly homogeneous throughout its entire extent.

Side-wall junction

From the doping concentration simulator we find that the metallurgical surface of the side-walls of the pn-junction are situated at approximately $5310\,nm$ away from the corresponding parallel plane passing through the centre of the device. Thus, our test pn-junction will be said to have 'real' dimensions of 10.62×10.62 microns. The doping profile simulator shows a doping gradient of approximately $\xi_{side} = 2.99 \times 10^{14}$ $dopants/cm^3/nm$ at a depth of $100\,nm$.

Using the same process as for the areal junction we compute $V_{bi} \approx 0.612\,V$ and $W \approx 9\,nm$, with zero-bias depletion region edge doping concentration of 1.35×10^{15} $dopants/cm^3$.

Under bias, the calculation from equation 4.100 yields a ratio of 2.093 and a depletion region width under bias equal to $18.8\,nm$. At that distance from the metallurgical surface, the net doping concentration is equal to approx. 2.82×10^{15} $dopants/cm^3$.Thus, the average doping concentration throughout the depletion region will be At that distance from the metallurgical surface, the net doping concentration is equal to approx. 2.09×10^{15} $dopants/cm^3$.

Similarly to the areal junction we shall consider that the side-wall is adequately represented by these numbers throughout its entire extent. We ignore side-wall corners where one side-wall meets another.

Depletion region summary

At the end of this process we have obtained the data populating table 4.6.

4.3.3 Simplifications

The idealised pn-junction model we have set up is still very complicated. Therefore it will be further simplified for the purposes of carrying out our worked example. Starting from the idea that we are mainly interested in fractional modulation depth we can make significant simplifications in how we regard doping concentration far away from the depletion region. For this reason we shall consider that the doping concentration 'inside' the N-well is uniform and always equal to 10^{19} $dopants/cm^3$ whilst the substrate will be assumed to be uniformly doped at 10^{16} $dopants/cm^3$. We shall thus, in essence consider

[10] Volume of semiconductor defined by the sweep of the depletion region boundary surface when the bias voltage across it changes from low to high state. Consists of two sub-volumes; one either side of the metallurgical surface.

Table 4.6: Summary of depletion region-related parameters computed throughout this example. Param.: parameter symbol. dop: dopants.

Param.	**Value in areal**	**Value in side-wall**	**Units**	**Description**
ξ	1.14×10^{14}	2.99×10^{14}	$\frac{dop}{cm^3 nm}$	Doping gradient
V_{bi}	0.578	0.612	V	Built-in voltage
D_{areal}	2530		nm	Areal junction depth
D_{fringe}	5310		nm	Side-wall junction distance from device centre
W_0	12.4	9.0	nm	Zero-bias depletion width
W_{-5}	26.4	18.8	nm	$5\,V$ reverse bias depletion width
R	2.129	2.093	-	W_0/W_{-5}
N_0	0.7×10^{15}	1.35×10^{15}	$\frac{dop}{cm^3}$	Net doping at distance $W_0/2$ from metallurgical surface
N_{-5}	1.5×10^{15}	2.82×10^{15}	$\frac{dop}{cm^3}$	Net doping at distance $W_{-5}/2$ from metallurgical surface
N_{avg}	1.1×10^{15}	2.09×10^{15}	$\frac{dop}{cm^3}$	Average of N_0 and N_{-5}

any point in the semiconductor volume that does not belong to a depletion region at any bias voltage as uniformly doped for simplicity.

Furthermore, we shall assume that the areal junction is perfectly perpendicular to both the side-wall junction components and the incoming illumination before treating the depletion regions of areal and side-wall junction components as homogeneous regions featuring the average doping concentration computed for their entire extents. Finally, we shall only consider the main beam component that is transmitted through each refractive surface from emitter to detector.

4.3.4 Absorption coefficients and model summary

Finally, we can add information on absorption electro-optical coefficients vs. doping and free carrier type from Figure 4.3 on page 112 to complete the set-up process.

Note: net doping concentrations either side of the junction will be the same by symmetry, but absorption coefficients will not because of the different nature of free carriers at those locations. This is important in the computation of absorptive losses through the areal and side-wall depletion region edge swept volumes. However, because the depletion regions swept volumes either side of the pn-junction are geometrically symmetrical and Beer-Lambert losses through two consecutive lossy media layers of equal width follow $e^{-a_1 \Delta x/2} e^{-a_2 \Delta x/2} = e^{-\frac{a_1+a_2}{2} \Delta x}$, we can simply take the average absorption coefficient either side of the junction, assign that to the whole swept volume and apply Beer-Lambert on both sides of the swept volume as if they were a single layer (the argument is fully elaborated upon in section 3.2.3).

At the end of the set-up process we can summarise the following key parametrs:

Table 4.7: Some key parameters used for the worked example in this section. A: illuminated Silicon area. D: Die front-end thickness. d: N-well metallurgical surface depth. w: metallurgical surface breadth in lateral direction (side length of our square junction). N_A: Substrate doping concentration ('acceptor' concentration). N_D: N-well doping concentration ('donor' concentration). n_i: Intrinsic carrier concentration. λ: Radiation wavelength. a_{sub}: Substrate absorption coefficient. a_{N-well}: N-well absorption coefficient. a_{areal}: Absorption coefficient in idealised/homogenised areal junction swept depletion volume. a_{fringe}: Absorption coefficient in idealised/homogenised side-wall junction swept depletion volume. a_{SiO2}: Silicon dioxide absorption coefficient. V_1: Bias voltage 1. V_2: Bias voltage 2. I_0: Irradiance at departure from light source. T: Absolute temperature.

PARAMETER	**VALUE**	**UNITS**
A	121	μm^2
D	300	μm
d	2.53	μm
w	10.62	μm
N_A	10^{16}	$dopants/cm^3$
N_D	10^{19}	$dopants/cm^3$
n_i	10^{10}	$dopants/cm^3$
λ	1550	nm
a_{sub}	2.4×10^{-6}	μm^{-1}
a_{N-well}	1.3×10^{-2}	μm^{-1}
a_{areal}	3.13×10^{-7}	μm^{-1}
a_{fringe}	5.49×10^{-7}	μm^{-1}
a_{SiO2}	approx. 0	μm^{-1}
V_1	0	V
V_2	-5	V
I_0	1	$\mu W/\mu m^2$
T	298	K

4.3.5 Computation of modulation

Combining information from tables 4.5 and 4.6, keeping the simplifications made under consideration and using fundamental electro-optical modulation theory from this chapter (section 4.1) we have enough information to compute modulation. We shall now further split the pn-junction into the following regions in order to improve the quality of our analysis:

- Component I: A.k.a. the areal junction. The beam enters uniformly doped N-well and first crosses the n-side depletion region swept volume. Next, it crosses the areal depletion region, then the p-side swept depletion region volume and finally exits die after crossing lightly doped substrate.
- Component II: Side-wall region 'strips' that will be depleted once the reverse-bias voltage is cranked up to $5\,V$, but are undepleted at zero-bias. In essence they form the swept volume of the periphery of our pn-junction and for all intents and purposes they are the functional part of the hitherto called 'side-wall junction' component. Light enters the swept volume that is simplified to be uniformly doped with the average depletion region doping concentration, then proceeds through substrate and finally exits the die. This component includes the swept volumes from either side of the pn-junction.
- Component III: Side-wall dead zone. This part of the side-wall junction is depleted in both states. The beam enters this region, crosses it unabated, then crosses the substrate and finally exits the die. This region plays no active role in modulation under our biasing regime, but affects the overall level of light reaching the photodetector.
- Component IV: Peri-Silicon. This region lies outside the device and everything influenced directly by the ebb and flow of the depletion regions but is illuminated nevertheless. It consists entirely of uniformly doped substrate. The beam enters this region and then exits the die.

Common losses:

Analysis will proceed independently for each component, but reflective losses at all the refractive interfaces encountered between emitter and detector will be common for all.

Using information from table 4.1 on page 111 we compute a main beam component transmittance of approx. $T_0 = 55.64\%$. This forms an upper bound in terms of the main component, on top of which additional losses through the free carrier-laden Silicon will be inflicted.

Component I:

We begin by calculating transmittance in component I. In this straightforward case the beam enters the Silicon into uniformly doped N-well territory.

Next, we need to calculate the path length through the N-well, which is equal to the n-side depletion region width under $5\,V$ reverse bias subtracted from the depth of the metallurgical surface. The resulting number is $\Delta x = 2530 - 13.2 = 2516.8\,nm$.

Subsequently, Beer-Lambert absorption yields a transmittance through the N-well equal to:

$$K_{I,N-well} = \frac{I_{out}}{I_{in}} = exp(-\alpha_{N-well}\Delta x) = 96.7811\% \tag{4.101}$$

where $K_{I,0}$ represents transmittance through Silicon for junction component I at the $N-well$ region, and I_x represents the irradiance at point x.

Carrying out the same calculations for the depleted region swept volumes (both identical in extent and consistency) we obtain: $\Delta x = 14\,nm$ and $K_{I,swept} = 99.9999(995618)\%$, where $K_{I,swept}$ represents transmittance through both sides of the depletion region swept volume when this region is not depleted and Δx represents the combined thickness of both n- and p-side swept volumes. The last few digits of $K_{I,swept}$ are in parentheses to indicate at what point the number becomes 'interesting' and starts carrying genuine information. Computations, however, will always be carried out to four significant digits as that convention has been used throughout the rest of the thesis (MATLAB default).

Finally, for the substrate we obtain: $\Delta x = 297483.2\,nm$ and $K_{I,sub} = 99.9286\%$.

We can now compute the overall transmittances through Silicon for component I at both bias voltages. At zero-bias:

$$\boxed{K_{I,0V} = K_{I,N-well} * K_{I,swept} * K_{I,sub} = 96.7120\%} \tag{4.102}$$

where $K_{I,0V}$ represents the overall transmittance for component I at $0\,V$ bias. Meanwhile, at $5\,V$ reverse bias be obtain:

$$\boxed{K_{I,-5V} = K_{I,N-well} * K_{I,swept} * K_{I,sub} = 96.7120\%} \tag{4.103}$$

we notice the results are identical to four significant digits.

Now we compute the total extent of the areal junction (component I) as $w^2 = 112.7844\,\mu m^2$ and we can finally aggregate all our information into the total transmitted power for each bias case. We obtain transmitted power of:

$$\boxed{P_{I,0V} = P_{I,-5V} = I_0 T_0 K_{I,0V} w^2 \approx 60.6900\,\mu W} \tag{4.104}$$

at both bias conditions, where $P_{I,xV}$ represents transmitted power for junction component I at $x\,V$ forward bias.

Since the results are so close to identical, we can not define a non-zero fractional modulation depth and we thus consider that the areal component is electro-optically, effectively inactive (transmitted power is not a function of bias voltage). In order to examine theoretically what sort of modulation we can expect from the system laid out above, we have decided to compute and quote $P_{I,0V}$ and $P_{I,-5V}$ to enough significant digits so that we notice some difference:

$$P_{I,0V} = 60.689912287\,\mu W \tag{4.105}$$

$$P_{I,-5V} = 60.689912553\,\mu W \tag{4.106}$$

This very rough calculation yields a fractional modulation depth of approx. $4.4\,ppb$.

Component II:

In this component we consider the strips of side-wall junction that are defined by the sweep area of the depletion region. We begin by noting that their length has been implicitly assumed to be equal to the depth of the metallurgical surface of the areal junction (we have ignored areal/side-wall transitional regions). This is equal to $2530\,nm$ and represents the path-length of light through the strip (our Δx).

Now we can combine this with the absorption coefficients under various bias conditions and calculate transmittance. For zero-bias, the region is undepleted and therefore features a non-zero 'effective' absorption coefficient, which from table 4.7 is equal to $5.49 \times 10^{-7}\ dopants/cm^3$. We thus compute the transmittance through the undepleted swept volumes of component II as:

$$K_{II,swept} = exp(-\alpha_{fringe}\Delta x) = 99.9998(611)\% \tag{4.107}$$

The rest of the free carrier absorption losses are through the substrate. We compute $\Delta x = 297470\,nm$ and transmittance $K_{II,sub} = 99.9286\%$.

Thus we obtain the following transmittances through Silicon for different bias conditions:

$$\boxed{K_{II,0V} = K_{II,swept}K_{II,sub} = 99.9285\%} \tag{4.108}$$

$$\boxed{K_{II,-5V} = K_{II,sub} = 99.9286\%} \tag{4.109}$$

Next, we compute the area that component II occupies and find it approximately equal to $4w(W_{-5} - W_0)$, i.e. the perimeter of the metallurgical surface of the side-wall junction component multiplied by the combined thickness of both swept volumes. This yields a value of $0.456\,\mu m^2$.

We can now compute total transmitted power through component II for different bias conditions:

$$\boxed{P_{II,0V} = I_0T_0K_{II,0V}[4w(W_{-5} - W_0)] = 0.2535\,\mu W} \tag{4.110}$$

$$\boxed{P_{II,-5V} = I_0T_0K_{II,-5V}[4w(W_{-5} - W_0)] = 0.2535\,\mu W} \tag{4.111}$$

Thus the fringe is thus, also to a certain approximation optically inactive, which means that the entire device under consideration is too weak to create any visible modulation at the current level of

accuracy. (Components III and IV do not change in response to bias voltage by definition). In order to examine theoretically what sort of modulation we can expect from the system aid out above, we have decided to compute and quote $P_{II,0V}$ and $P_{II,-5V}$ to enough significant digits so that we notice some difference:

$$P_{II,0V} = 0.25353699\,\mu W \tag{4.112}$$

$$P_{II,-5V} = 0.25353724\,\mu W \tag{4.113}$$

This very rough calculation yields a fractional modulation depth of approx. 986 *ppb*. This is considerably stronger than the equivalent metric for the component I (areal junction); a fact we can attribute to the significantly longer path length of radiation through electro-optically active regions of component II. Note: the side-wall may be a far stronger modulator, but it is also much smaller in extent than the areal junction. Thus the device as a whole will likely perform closer to the single-digit *ppb* range of the areal component than the equivalent metric for the side-wall.

Component III:

This component is different from its predecessors in that biasing does not affect transmittance through it. The first part of the light beam path length through this component concerns an absorption-free leg through permanently depleted Silicon. This extends for $2530\,nm$; the nominal depth of the areal component of the pn-junction.

The remaining $297470\,nm$ of path length through the die consist of substrate and we have already computed the transmittance for this case as $K_{III,sub} = K_{II,sub} = 99.9285\%$. This is also the overall transmittance in Silicon of component III, K_{III}.

We now compute the area of component III, which is approximately equal to $4wD_{areal}$, i.e. the perimeter of the metallurgical surface of the side-wall junction component multiplied by the total depletion region width at zero-bias, when the depletion region is at its narrowest. This is approximately equal to $0.418\,\mu m^2$.

We can therefore compute the transmitted power through component III:

$$\boxed{P_{III} = I_0 T_0 K_{III}[4wD_{areal}]} = 0.2324\,\mu W \tag{4.114}$$

Note: transmitted power lacks a biasing index and therefore only appears with the index indicating the component to which it corresponds.

Component IV:

The final component is the simplest to study. It concerns illuminated Silicon that does not form a functional part of the device, i.e. it is 'dead weight'. It consists of 300 microns of uninterrupted substrate

and we can calculate transmittance through a single application of the Beer-Lambert law with the substrate absorption coefficient. The result is transmittance $K_{IV} = 99.9280\%$.

The total area of component IV was defined as the difference between the illuminated 11×11 micron square and the other three components. It can be easily shown that the area covered by components I-III is a square with a side of $10.6388\,\mu m$, i.e. $\approx 113.1841\,\mu m^2$. Thus, the final component of the junction covers a total area of $A_{remnant} \approx 7.8160\,\mu m^2$.

We can now compute transmitted power through component IV as:

$$\boxed{P_{IV} = I_0 T_0 K_{IV} A_{remnant}} = 4.3457\,\mu W \tag{4.115}$$

Modulation throughout whole device - conclusions and summary:

In summary we have found that regardless of bias voltage, the device receives approx. $65.5216\,\mu W$ out of the $121\,\mu W$ being radiated towards it, with no modulation detectable at the ppm level. Given that the component responsible for most of the power transmission (I) shows single-digit ppb modulation levels it is reasonable to expect the whole device to exhibit modulation in the ppb range, as we have seen. This dramatic underperformance in comparison to our abrupt junction example (section 4.2) is largely attributable to the fact that sweep volumes of nm-scale widths (similar to their abrupt junction example counterparts) now contain dramatically smaller concentrations of dopants. Where previously sweep volumes contained doping concentrations in the 10^{17} and 10^{20} $dopants/cm^3$ orders of magnitude, now they are reduced to containing 10^{15} $dopants/cm^3$.

The final conclusion of this endeavour is that electro-optical modulation in CMOS Silicon may turn out to be of extraordinary weakness. This, in turn, may mean that a solid, engineerable and useful electro-optical modulation platform on the basis of the free carrier absorption phenomenon and the geometrical configuration of our set-up (i.e. without involving waveguides and other very convenient and expensive post-processing) may prove elusive. However, this must be combined with the fact that improvements in the system used to drive the modulators and detect the resulting modulation can potentially improve performance dramatically compared to our worked examples. As a simple example, working in mid-infra-red (MIR) wavelengths of e.g. $5 - 8\,\mu m$ can potentially improve modulation by a factor of roughly $10 - 27$. For another example let us consider that increasing reverse bias to $10\,V$ could potentially improve modulation by a factor of over 2 vs. a system that operates a $2\,V$ reverse bias voltages [11].

Of course, results from our worked examples must be understood within the context of errors. When the numbers involved are so small, errors are bound to intervene. Constants, such as Euler's number were used with a large number of significant digits, but uncertainties such as the one arriving from the calculation of the absorption coefficients of free carriers for given doping concentrations could not be avoided and are expected to change the results quite significantly, and yet maintain them in the same 'ballpark'. At the end of the day, the only way to find out what really happens is through experiment.

[11] Under the simplifying assumption that the junction is abrupt and modulation scales with the square root of the bias voltage. For linearly graded junctions the benefit is smaller as we have seen the scaling to be with the cubic root of bias voltage.

4.4 Conclusions

In this chapter we have performed a theoretical study of realistic modulator configurations based on realistic doping profiles.

Importantly, a basic analysis of optical amplitude modulators relying on the phenomenon of free-carrier absorption in realistic devices were carried out. For the latter case expressions that can help assess system performance have been derived.

Furthermore, a worked example based on a simple, abrupt junction case was given and the relations between absolute modulation depth and various key parameters such as voltage biasing, device geometry and light source irradiance were specified.

It is important to note that the theory explained throughout this chapter can be expanded to a much greater degree; perhaps as much as developing it into a tool that would use numerical solvers and finite element analysis in order to yield estimates of modulator performance. Nevertheless such advanced analysis lies beyond the scope of this project.

Bibliography

[1] R. Kitamura, L. Pilon, and M. Jonasz, "Optical constants of silica glass from extreme ultraviolet to far infrared at near room temperature," *Appl. Opt.*, vol. 46, pp. 8118–8133, Nov 2007.

[2] D. Schroder, N. Thomas, and J. Swartz, "Free carrier absorption in silicon.," *IEEE journal on solid state circuits*, vol. SC13, pp. 180–187, 1978.

[3] R. A. Soref and B. R. Bennett, "Electrooptical effects in silicon," *IEEE journal of quantum electronics*, vol. QE-23, no. 1, pp. 123–129, 1987.

[4] C. C. Katsidis and D. I. Siapkas, "General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference," *Appl. Opt.*, vol. 41, pp. 3978–3987, Jul 2002.

[5] A. Serb, K. Nikolic, and T. G. Constandinou, "A CMOS-based light modulator for contactless data transfer: theory and concept," *SPIE proceedings*, vol. 7943, pp. 794317–794317–12, 2011.

[6] H. C. Huang, S. Yee, and M. Soma, "Quantum calculations of the change of refractive index due to free carriers in silicon with nonparabolic band structure," *Journal of Applied Physics*, vol. 67, pp. 2033 –2039, feb 1990.

[7] S. M. Sze, *Physics of Semiconductor Devices*. Wiley, 2nd ed., 1981.

Chapter 5

CMOS electro-optical modulator design and results

The mandate of this thesis included a single main objective: investigate electro-optical phenomena in Silicon with the aim of enabling the creation of a communications platform for contactless CMOS integrated circuits. This was split into three components, namely developing the data read-out module, the data read-in module and the power scavenging and management system. In previous chapters we have examined the relevant literature and laid out the theoretical groundwork for the realisation of these component objectives. This included a study of the phenomena that can be engineered in order to allow the function of such communication system and some theoretical tools that would help us gain an understanding of how the practical implementation of devices that exploit the said phenomena can be carried out and estimate their performance.

In this chapter we tackle one of the objective components, namely the data read-out module. This includes a brief overview of the modulator structure designs used throughout the project (the full design repository can be found in the appendix A) and results obtained from the testing and measurement protocols applied to them. Information describing the test-bench set-up and operation in detail is included in this chapter. Raw data measurements are also found in the appendix, leaving only processed results within this chapter (appendix B). Any trends with regards to modulator performance vs. technological node, junction type, device geometry etc. will also be part of the result interpretation section. The same applies for the interpretation of our results from the point of view of the prospects of practical implementation of these devices as parts of a full opto-electronic communications system.

This chapter is organised as follows: In section 5.1 we define the objectives that each modulator device must fulfill, lay out the constraints imposed by the project specifications and proceed to examine the 'three levers' that behoove engineering, determining likely outcomes resulting from their manipulation. In section 5.2 an overview of our design choices, including the rationale behind them is presented. The section also includes a handy table summarising all device designs. Next, section 5.3 showcases preliminary results from the proof-of-concept phase of the project. Section 5.4 details the construction and operation of the test-bench used for our main investigation. Conclusions pertaining to comparisons between devices will be tackled in section 5.5, as will information regarding the practicality of implementing any of the designed devices as modulators in a realistic setting. Section 5.6 will feature a discussion on additional considerations, such as the effects of various sources of error and design recommendations

for future work.

5.1 CMOS optical modulators: aims, constraints and design considerations

Before any design action can be taken, both the aims of the design and the considerations they imply need to be clear. All devices created during this research project had as their primary aim the ability to function as near infra-red (NIR: 1.1 - 1.8 micron wavelengths) modulators[1] and provide valuable insight into how various design choices translate into modulator performance. This is in stark contrast to the aims of typical industrial applications where the objective concerns the manufacturing of optimal devices. The aforementioned aim almost completely overpowered any secondary aims (such as, for example, the ability of the very same devices to harvest optical power from the visible part of the EM spectrum).

A caveat was also imposed that all manufacturing had to be carried out in commercially available CMOS technology without the plethora of complicated, additional manufacturing steps that are invariably employed to create waveguides on Silicon. This constraint removed a large number of available, but generally niche processing options, such as the creation of complicated, circular structures that allow for the implementation of reasonably high quality Silicon waveguide micro-resonators. As a result, the possibility of creating planar waveguides and modulating light in a plane parallel to the surface of the die was excluded and the approach described in 4.1.1 was adopted instead. Modulator devices therefore took the form of simple, CMOS pn-junctions. This generated the following design considerations: a) selecting a suitable manufacturing process, b) choosing a pn-junction type and c) defining the geometry of each device.

Beside the restrictions imposed by the project specification, numerous practical restrictions stemming from such diverse factors as availability of suitable components, instrumentation etc. had to also be taken into consideration, however these will be presented in the section on the design and operation of the test bench (5.4.2).

5.1.1 Manufacturing technology

The selection of technology is a process that has to take into consideration not only performance aspects, such as estimated modulation performance and the capability to sustain high-quality analogue circuitry, but also practical aspects such as the availability of departmental experience and design tools for the technology.

In terms of expected performance, different technologies will feature pn-junctions with different doping profiles. Specifically technologies of lower node (smaller feature sizes) tend to grant access to higher doping concentrations and steeper dopant concentration gradients overall and lead to improved modulation efficiency[2]. In order to probe for that effect, tests were conducted in technologies representing three different processing nodes.

[1] This range of operation is chosen for reasons already explained in section 4.1.1.

[2] Note: Because of the secrecy veiling doping profiles, very generic comparative results of doping concentrations can only be extracted experimentally or estimated by looking at device electrical characteristics (diode forward bias voltages for example).

Furthermore, considerations from the power recovery system also influenced the decision made on technology selection since some analogue circuitry had to be implemented in order to handle power harvesting and management. Thus, well-established, 'analogue-friendly' technologies of higher technological nodes were considered as the best candidates.

In the end a triplet of technologies was selected in the form of the AMS 0.35 micron C35 process (hence forth: AMS35), the IBM 0.18 micron H18 7HV process (hence forth: IBM18) and the UMC 0.13 micron UMC13 process (hence forth: UMC13). These were chosen as a good representative selection of technologies that can handle both analogue and digital circuitry and also demonstrate the differences between technologies with different doping profiles. Key parameters of these technologies were available to us only via non-disclosure agreements (NDA), and for that reason publishing relevant data is not possible.

5.1.2 pn-junction type

In section 3.3 we have seen how diffusion, implantation or combinations of both can be used to create pn-junctions. Drawing from theoretical results (derived and shown in chapter 3 and also alluded to in section 4.2.4) we expected to find that devices that feature high doping concentrations at the volumes enclosed within the border surfaces of their depletion regions at some minimum and maximum bias voltages will be better modulators. As such, a good approach would be to identify and choose junction types that offer these high doping concentrations at those strategically important locations.

Matters, however, are complicated by the highly heterogeneous structure of realistic pn-junctions. The simple theoretical results of doping profiles resulting from simple doping processes shown in 3.3 cannot encompass the entire complexity of the said doping profiles, but do make an important distinction into the following types of pn-junctions: a) diffusion-based, b) base (areal) well-based and c) fringe (perimetric) well-based junctions.

In more realistic CMOS pn-junctions, more advanced terminology needs to be defined in order to describe the said junctions in sufficient detail. Throughout this chapter we shall use such terminology as defined in Figure 5.1.

The archetypal diffusion-based junctions created by the process of planar diffusion will consist of a base area (away from the edges) where the doping profile vs. depth is a very weak function of coordinates on the die surface and the edges where edge/corner effects will become significant. These edge effects have not been considered in the theory section of this thesis yet merit an empirical study. As such the design of devices based on diffusion junctions with different area/perimeter ratio was implemented as a means of providing an empirical, qualitative relation between so-called 'fringe' and 'areal' junctions.

Well junctions are even from the point of view of the simplified theoretical analysis of section 3.3 split between separate basal and fringe domains, the former being determined primarily by the Gaussian, 'main', longitudinal dopant distribution and the latter by the lateral straggle effect (see aforementioned theoretical section for details). In reality well junctions are formed by not only both side-wall and areal regions, but also by the edge and corner regions between them; another subject for numerical analysis. As a result, the design of devices with different area/perimeter ratios was chosen as a means to resolving the contributions of each individual element, just like in the case of diffusion junctions. Note: side-wall and

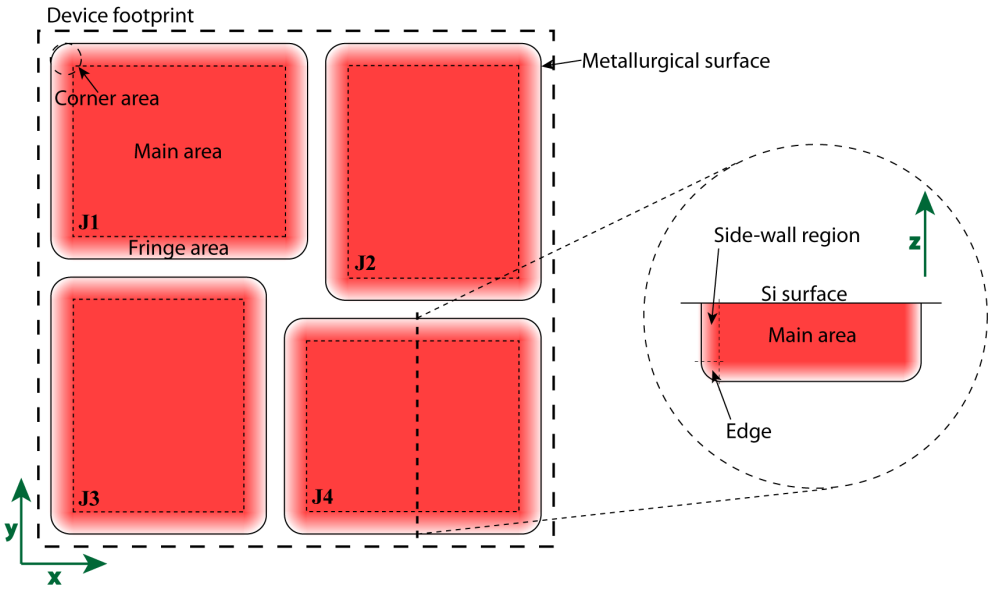

Figure 5.1: Technical terms used to describe devices, junctions and regions within junctions. The inset shows a cross-section of J4 at the location denoted by the dashed line bisecting J4. Separate axis sets are also given for the main image (x,y) and the inset (z) for clarity. Device footprint: the nominal size of the device. JX: Junction X within the device under study. Metallurgical surface: the nominal edge of a junction. Main area: the areal junction. Fringe area: regions close to the metallurgical surface of a junction in (x,y) terms. Corner area: regions close to the metallurgical surface corners of a junction in (x,y) terms. Side-wall junction (well-type junctions only): segment of the junction where lateral straggle is the dominant effect. Edge region: sectors of a junction close to the metallurgical surface in (x,z) or (y,z) terms excluding corners. Si surface: the interface where the front- and back-ends of the die meet.

fringe areas are not entirely interchangeable in this chapter as side-wall regions will denote well-defined vertical junction areas while fringe areas will denote the combination of side-wall and side-wall/base edge areas, the latter being regions where side-walls meet the areal junction region. In practice, we shall often use them interchangeably with the meaning of 'fringe' whilst in subsequent chapters this restriction is lifted entirely since the study of modulator and power scavenger performance as a function of geometry is simplified into the study of areal and fringe areas.

In conclusion, acknowledging the full complexity of real CMOS pn-junctions means that the objective of the study was oriented more towards splitting pn-junctions in functional and engineerable domains such as base and fringe areas rather than the theoretically defined base, side-wall, edge and corner regions and throughout the process determine what is the best methodology for designing practical optical modulators.

5.1.3 Device geometry

The topic of device geometry was largely addressed by the need to understand how different regions of junctions manufactured by use of different processes affect overall device performance, yet the matter of overall size still remained open. Large, square-footprint devices featuring sides spanning over 100 microns were chosen in the end for a number of reasons.

The main performance-related reason for choosing such large devices concerns issues of result quality. A large device can be used to host a couple of important design pattern families that we shall call 'coarse-grain' and 'fine-grain' patterns and helps capitalise on both their intrinsic advantages.

In coarse grain design, each device is split into relatively few, very large, physically independent (even though electrically connected) pn-junctions. The advantage of this approach is the option to minimise the prevalence of corner and edge regions throughout the device. This applies because the area of the designed device scales with the square of the footprint side length while edge areas scale linearly with the footprint side length. The properties of areal junctions can thus be studied with good accuracy. Meanwhile, the side-wall junction also scales linearly with the footprint side length of the device which means that side-wall and side-wall to base edge areas are linked to each other in a one-to-one relationship and can therefore be automatically treated as single 'fringe' regions within the junction. However, by using large footprint side length for our devices we can potentially reduce the prevalence of corners in our devices and thus obtain a clean estimate of fringe regions as a whole. Nevertheless, if the geometry of the pn-junction deliberately includes many corners, then the quality of the estimates for side-wall junction performance deteriorates in the same way that the quality of estimates for areal junction performance deteriorates in designs where many edge regions are present.

In fine grain design each device is split into a very large number of small, physically independent pn-junctions. The advantage of this design lies in that the small 'basic cells' that constitute the device can be characterised rather well due to their absolutely common design and the advantage of averaging results over a large number of basic cells. Thus, creating basic cells with slightly different geometries can potentially allow us to determine differences between subtle geometrical features. As an example one can study to what extent the presence of corner regions affects fringe regions by creating a rectangular basic cell as a reference and then comparing it to various 'trombone-extended' versions[3] (see Figure 5.2).

[3]The so-called 'trombone extension' technique refers to a design technique whereby a cell (in a 2D layout space) is cut

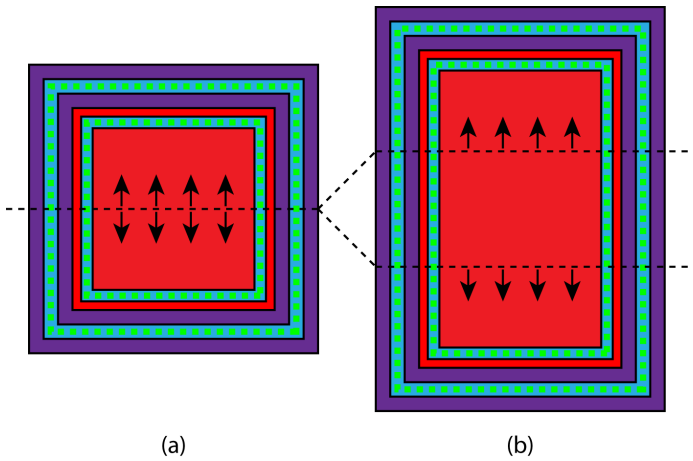

Figure 5.2: Trombone extending a physical device. a) Basic, unexpanded device. b) Expanded version. The dashed lines mark the borders of the added layout area. Legend: Purple - N-well, Red - Diffusion, Blue - Diffusive contact, Green (squares) - diffusion to metal 1 contact vias. Arrows indicate the 'pulling apart' action.

This is in contrast to the less subtle differentiations that can be made by coarse-grain devices.

Meanwhile, the main practical reason for choosing large devices pertained to alignment issues as aligning larger components was found to be easier and require less accurate positioner stages. More details about the practicalities of testing will be described in the appropriate section.

Finally, the square nature of the devices was chosen for practical reasons since it offers the advantage of having central symmetry. In terms of layout this makes for more flexibility in floorplanning as any square can be rotated by any integer multiple of 90 degrees and it will still fit within its original footprint while rectangles only show that property for integer multiples of 180 degrees.

5.2 Designing optical modulators in CMOS technology

After the aims and constraints of the project were determined and the design considerations taken into account the next stage of the optical modulator design cycle could begin: the design.

The necessity that our test devices should be designed in three different technologies and feature large, square footprints covering pn-junctions of different area/perimeter ratios representing both diffusion- and well-type 'flavours' was reflected in the task of designing optical modulators. In total four die designs were created and fabricated. The AMS35 and UMC13 technological processes were each represented by a single die design while the IBM18 technology was represented by two designs. Each design was given a unique code name: a) AMS35 - Ninja, b) IBM18 first attempt - Bean, c) IBM18 second attempt - Svejk, d) UMC13 - Teddy. The 'Bean' failed catastrophically due to a fault in the design of the pad ring and will not be discussed any further throughout this thesis. The other three dies all yielded results. Crucially, each technology was characterised by a unique die thickness. The UMC13 technology featured dies 279 μm thick, the H18 technology offered dies of 250 μm thickness and the C35 technology worked

in two constituent parts by means of a straight line, the constituent parts are then taken apart in a direction perpendicular to the bisecting line and the one-dimensional 'image' of the device along the section is extended into 2D between the inside edges of the trombone-extended cell.

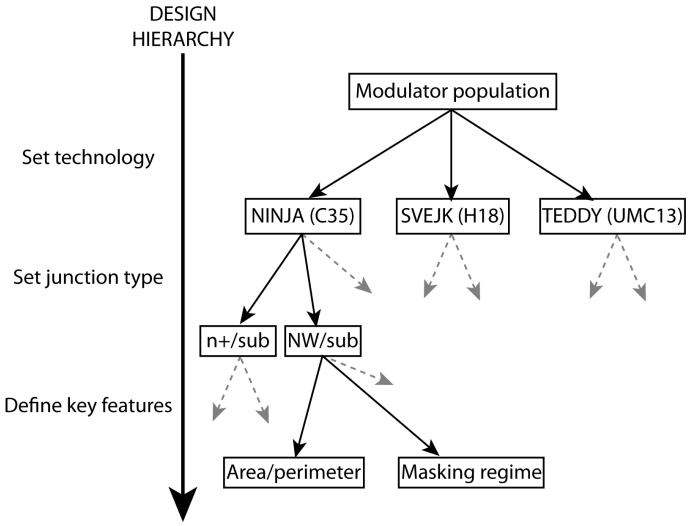

Figure 5.3: Design hierarchy for the entire modulator population generated throughout this project.

with thicker dies at $530\,\mu m$. Note: by comparison depletion regions in typical CMOS junctions are expected to be measured in nanometers (using physics from [1] and typical doping numbers found in CMOS).

All CMOS technologies offer even at their most basic at least three 'types' of pn-junctions: N-well to substrate (NW/sub), p-diffusion on N-well (p+/NW) and n-diffusion on substrate (n+/sub). Crucially, all CMOS technologies also allow the possibility of nested junctions. The simplest such example consists of p+/NW junctions. The NW sector of the junction sits on p-type substrate (here we consider CMOS and BiCMOS technologies that generally tend to reside on p-type substrates -see 3.3 for more information-). These junction types are assumed to work very differently from one another due to the distinct manufacturing processes employed in their formation, but also very differently from other junctions of the same type in different technologies for practically the same reason[4]. Meanwhile the behaviour of each of these junction families is expected to conform to certain, relatively simple rules that link modulator performance to geometrical features, specifically side-wall and areal elements of the junction. The combination of all these factors means that the junction type is a good way of categorising devices within a die before moving to further details with respect to geometry and other special features.

These design considerations naturally hint towards a design hierarchy that is based on technology, junction type and special geometrical (or any other type of) features; in that order. Thus, the design hierarchy can be illustrated as in Figure 5.3.

In order to quickly differentiate between devices and junctions the following convention will be followed throughout the rest of the thesis including the appendix: Each device on each die will be assigned a unique identifier number (Z) followed by a pair of letters representing the constituent parts of the junction (AB). This denomination will be preceded by the three letters (XXX) describing the die family and when discussing about physical dies as opposed to simply their designs a number identifying the die 'instance' (Y). This yields a code of the form:

[4]In the former case because of different procedures employed within a technology in order to create distinct types of doped regions and in the latter case because of different procedures employed across technologies to create the same doped regions.

$$\boxed{XXXY - Z - AB} \tag{5.1}$$

Thus, for example, the NW/sub segment of the 5th junction of the 3rd 'Teddy'-type device will be denoted as 'TED3-5-NS' where 'TED' stands for 'Teddy', 'N' for 'N-well' and 'S' for 'substrate'. When all objects of a certain type are to be selected, the corresponding part fo the code will be missing. Examples: a) The design of the 2nd device of any Teddy die will be marked as 'TED-2' ('Y' missing). b) All sub-junctions of device number 4 of the first Teddy die are marked as TED1-4 ('AB' missing). c) All devices on the 2nd Svejk die will be marked as SVJ2 ('Y' and 'AB' missing). Constituent parts of junctions are quickly summarised in table 5.1.

Table 5.1: Junction constituent part codes.

Code	**Part**
p	P-diffusion
n	N-diffusion
3	Triple well
N	N-well
S	Substrate

Note: even though codes of the form XXXY-Z-AA will be used throughout the thesis, full device descriptions will also be given throughout all main chapters for convenience. The appendix, however, will make almost exclusive use of the code notation.

A remark that applies to all devices on all dies concerns modularity. All devices are larger than $100\mu m \times 100\mu m$, which is over one hundred times the feature size. This alone qualifies them as large devices and due to this size a modular approach has been employed whereby each device consists of a collection of 'basic cells'. Each basic cell is a set of either one or more pn-junctions arranged in a geometric pattern that repeats itself periodically. All basic cells are connected in parallel. In nested junction devices, each junction forms its own 'basic cell'. Thus a p+/N-well/sub type device will feature p+/N-well and N-well/sub basic cells, each with their own characteristics and statistics. Nevertheless, in certain nested junction devices, there is a one-to-one correspondence between basic cells originating from different junction types. In those cases we refer to both those basic cells together as forming a 'device basic cell', which we will abbreviate to 'basic cell' for convenience.

On a larger scale, each device can also be thought of as an individual 'module' where the optically active part of the design is encircled by an electrical biasing guard-ring intended to ensure that the substrate all around the device is properly grounded. 'Off-shoots' (wires springing from the guard ring) criss-cross our active device placing substrate anchors at specified intervals in order to ensure that no part of any device is too far from a grounding point. Unfortunately, our designed devices are not planar and as such cannot by nature be easily isolated from the substrate (particularly N-well based structures can never be isolated in the technologies we chose to work with). On the other hand, ESD (electro-static discharge) guard-rings (diode-based) have not been explicitly placed except where the pad-ring includes them by default as our diodes: a) were handled under controlled conditions with ESD protection being worn at all times during handling and therefore ESD issues never arose and b) being very large in area we expected them to be highly resistant to ESD shocks.

Table 5.2 summarises the modulator designs developed throughout this project although the designs in full detail are 'stored' in the design repository in the appendix, chapter A.

5.3 Early experiments and results

Historically, the progression of the project can be divided into two sections: a) An early stage where a very simple test bench was assembled for proof-of-concept testing and b) a late stage where the full characterisation of the observed electro-optical modulation effects was to be carried out. This section contains information on the early stage, proof-of-concept experiments, much of which has already been published in [2, 3]. We shall begin by providing a very brief overview of the test bench before proceeding to describe some preliminary results.

5.3.1 Proof of concept test bench

The test bench used for our initial experiments was similar in spirit to what was used for the main experiments. It consisted of a light source, a test die hosting modulator devices (in this case a C35 sample, the same as the C35 dies described in the design repository found in the appendix, section A) and a photodetector with associated instrumentation, tasked with probing for modulation.

The light source was an off-the-shelf infra-red LED with central wavelength at $1550\,nm$ (ThorLabs - LED1550E), connected to a source-meter unit that provided a constant current bias (Keithley - Source-meter unit 6430). The modulators were manually connected to a signal generator via an on-PCB switch-board. The photodetector was then connected to a crude, in-house-built transimpedance amplifier, which in turn fed into a lock-in amplifier. The output of the lock-in amplifier was finally routed to an oscilloscope. Under ideal conditions the light source, modulator and photodetector constituent parts of the set-up would be completely independent from an electrical point of view; that is to say, no signal feed-through or pick-up from any constituent part should contaminate any other constituent part. An overview of the test-bench is shown in Figure 5.4.

The core of the test-bench consisted of the LED, the test chip and the photodetector. These components had to communicate with each other in a purely optical fashion. This created the issue of positioning them within an assembly that ensures their proper alignment at all given times and precludes any optical contamination from outside. This set of constraints let to our proposed, PCB-based solution: The LED, test chip and photodetector were physically entombed inside a stack of PCBs crafted in such way as to create a cavity within which light could exit the LED, travel through the test chip and hit the photodetector with minimal outside interference (see Figure 5.4 parts (b) and (c)).

The PCB stack consisted of 11 layers of which five were functional and the remaining six were simply spacer cells meant to maintain an appropriate distance between the functional PCBs. All PCB layers featured three pairs of holes of which two were used to bolt the entire assembly together and the last pair was used for the insertion of alignment pins, meant to maintain the alignment of the PCB stack. Finally, gold sheets on each PCB and vias ensured that all PCBs shared a common ground plane.

In more detail, the functional layers of the PCB stack were:

Table 5.2: Summary of all devices designed within the framework of this project. The A/P ratio expresses the exposed area over exposed perimeter ratio. Special features are an allusion to any other features aside from geometry used to tailor each individual device. In multiple junction devices the sub-junction in parentheses denotes the key sub-junction of the device. A/P data corresponds to that sub-junction. The 'perim mask' label indicates that a mask covering the perimeter of some sub-junction has been used. This is explicitly given except in cases where the answer is obvious.

MODULATOR DESIGN SUMMARY

Device ID	**Junction type**	**Group**	**A/P** (μm)	**Footprint** (μm^2)	**Special feature**
NIN-1	NW/sub	-	1.208	479×479	
NIN-2	(p+/NW)/sub	-	0.183	479×479	
NIN-2	p+/(NW/sub)	-	7.317	479×479	
NIN-4	n+/sub	-	16.742	479×479	
NIN-5	NW/sub	-	3.660	479×479	
NIN-6	NW/sub	-	3.660	479×479	No passivation
NIN-7	(n+/p+)/NW	-	0.600	479×479	Butting
NIN-8	n+/p+	-	0.251	479×479	
NIN-9	n+/sub	-	0.259	479×479	Poly
SVJ-1	n+/sub	G1	∞	300×300	Perim mask
SVJ-2	n+/sub	G1	∞	300×300	Perim mask
SVJ-3	(p+/NW)/sub	G2	1.816	300×300	NW perim mask
SVJ-3	p+/(NW/sub)	-	∞	300×300	NW perim mask
SVJ-4	(p+/NW)/sub	G2	∞	300×300	NW & p+ perim mask
SVJ-4	p+/(NW/sub)	-	∞	300×300	NW & p+ perim mask
SVJ-5	n+/sub	G3	1.562	300×300	
SVJ-6	n+/sub	G3	∞	300×300	Perim mask
SVJ-7	NW/sub	G4	0.316	200×200	
SVJ-8	NW/sub	G4	0.455	200×200	Rectangular cell
SVJ-9	NW/sub	G4	0.973	200×200	
SVJ-10	NW/sub	G4	0.973	200×200	Rectangular cell
SVJ-11	NW/sub	G4	2.567	200×200	
SVJ-12	(3W/NW)/sub	G5	0.596	200×200	
SVJ-12	3W/(NW/sub)	-	5.882	200×200	
TED-1	NW/sub	H1	0.264	495×495	
TED-2	n+/sub	H2	0.285	495×495	
TED-3	n+/sub	H2	9.313	495×495	
TED-4	NW/sub	H1	19.369	495×495	
TED-5	(3W/NW)/sub	H3	0.342	495×495	
TED-5	3W/(NW/sub)	-	10.979	495×495	
TED-6	n+/p+/sub	H5	0.320	299×299	Butting
TED-7	(n+/3W)/NW/sub	H4	6.827	299×299	
TED-7	n+/(3W/NW)/sub	-	7.885	299×299	
TED-7	n+/3W/(NW/sub)	-	8.924	299×299	
TED-8	(n+/3W)/NW/sub	H4	1.219	299×299	
TED-8	n+/(3W/NW)/sub	-	1.919	299×299	
TED-8	n+/3W/(NW/sub)	-	9.356	299×299	
TED-9	(3W/NW)/sub	H3	1.935	299×299	
TED-9	3W/(NW/sub)	-	9.579	299×299	

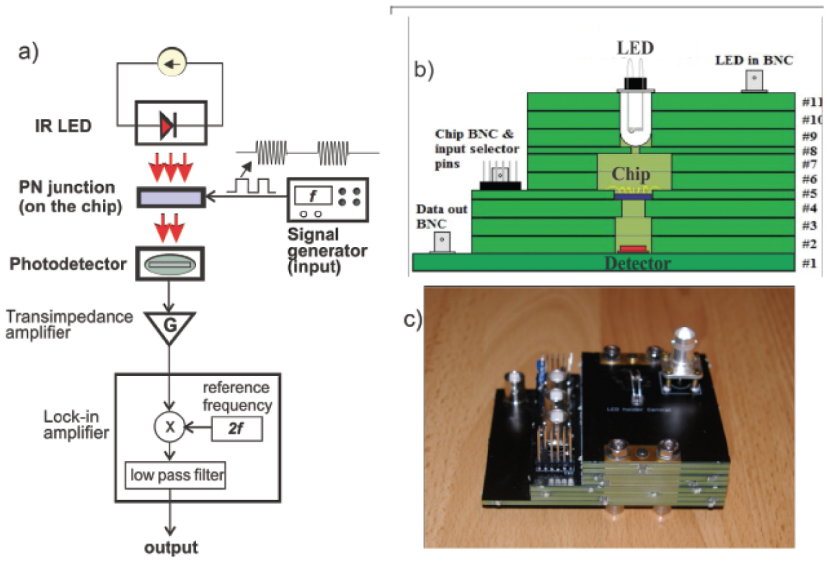

Figure 5.4: Test bench for proof-of-concept phase experiments. Reproduced from our own publication [2]. a) Basic diagram showing electrical connections and optical path (red arrows). b) Cross-section of the PCB stack within which our LED, test chip and photodetector were entombed. c) Picture of the PCB stack set-up.

- #1: The photodetector and the BNC port linking it to the outside world were hosted on this layer.
- #4: The test die was mounted on top of this PCB. The test die had to be mounted in such way as to allow light to pass through it. As such, it was decided to mount the die exactly over a hole slightly smaller in diameter than the diagonal of the square test die we were using at the time. The die was mounted by its corners on a gold sheet that surrounded the hole. The gold sheet was electrically connected to GND.
- #5: This die received the bond wires from the test die and hosted the switchboard via which test devices were connected to the signal generator.
- #8: This layer was intended as an optical mask. The PCBs intended for use as layer 8 components featured $0.5\,mm$ diameter holes positioned in such way as to allow illumination of a single device on the test die. Eight sets of layer 8 'pin-hole' PCBs were manufactured: one for each test device.
- #11: This layer hosted the LED and the port linking it to its bias current source.

This test-bench was supposed to work very well in theory, but in practice a number of opto-mechanical issues cropped up:

- The perfect alignment of all PCB layers was never achieved fully.
- Mounting the die on its layer 4 host PCB solely by its corners was a very hard task in practice.
- The layer 8 masks were never manufactured with sufficient pin-hole diameter control and as such were quickly rendered useless.

- The set-up never offered a way for changing the alignment between the light source and the test-die or photodetector components. Thus, any non-uniformity in the beam profile arriving from the LED remained uncompensated.
- Changing test dies involved disassembling the stack and reassembling it with a different set of layer 4 and 5 PCBs.

The operating protocol of the test bench was as follows:

- Light source: Either bias with such constant current as to maintain an optical output power level of approx. $2\,mW$ or keep shut down.
- Set signal generator to provide 'bursts' of $1\,kHz$ sinusoidal stimuli between $0\,V$ and $2\,V$ reverse bias at a burst frequency of $1-10\,Hz$ with $50\,\%$ duty cycle ratio (i.e. 0.5 seconds of $1\,kHz$ stimulation followed by 0.5 seconds of silence).
- The lock-in amplifier was set to detect the second harmonic at $2\,kHz$ and the phase of its internal reference sinusoid was set such as to maximise output magnitude[5]. This magnitude was considered to be the 'signal strength' corresponding to the device under test at the given illumination and biasing conditions. The integration time for the lock-in amplifier was set at $3\,ms$.

When in the dark, the lock-in amplifier would typically lock on to pick-up feeding from the signal generator, through the modulator die and into the photodetector whilst under illumination pick up plus optical modulation would be detected concurrently. Given the information available at the time, it was considered that the absolute difference between the strength of the signals arriving from the lock-in amplifier under illumination and in the dark would represent the effects of electro-optical modulation on top of the pick-up:

$$M \propto |S_{light} - S_{dark}| \tag{5.2}$$

where M represents modulation and S_x represents signal strength under conditions x. The quantity at the right hand side of the equation was measured in mV.

Note: The reason why the lock-in amplifier was set to detect activity at the second harmonic was because pick-up at the first harmonic was extremely strong. The signal generator was constructed to feature very low harmonic distortion but the modulator, being a pn-junction, was a highly non-linear device. Specifically, since depletion region width varies approximately with the square of the externally applied bias voltage we expect that stimulation with a pure sinusoid will create a response featuring a large second harmonic component (see [2] for more details). Since modulation depends exponentially on

[5]The lock-in amplifier operates by multiplying an incoming signal with an internally generated reference sinusoid signal and integrating over a certain period of time. The details are provided in full in the section describing the test-bench for our full characterisation experiments (section 5.4) but the net effect is that the output of the lock-in amplifier will be a DC signal that depends on the intensity of the lock-in frequency component of the input signal and the phase between the input signal lock-in component and the reference sinusoid.

depletion region width, it was considered that by seeking our signal at the second harmonic we would be avoiding at least a large component of direct pick-up from the signal generator whilst still running a high chance of observing modulation.

5.3.2 Results

Despite all the issues of our rudimentary test-bench, a set of interesting results was collected. It was found that indeed the lock-in amplifier indicated that the $2\,kHz$ component it was detecting did have different magnitudes depending on whether the LED was active or not. Extracted $|S_{light} - S_{dark}|$ quantities are summarised in table 5.3.

Table 5.3: Results of our preliminary, proof-of-concept stage tests. Left column: device identifier. Middle column: device type (see table 5.1 on page 165). Right column: measure of electro-optical modulation as a result of the application of the same biasing and illumination protocol to each test devices ($|S_{light} - S_{dark}|$).

DEV	**TYPE**	**Net modulation**
		mV
1	NS	290
2a	NS	19
2b	pN	18
4	NS	12
5	NS	235
6	NS	350
9	nS	120

For reasons that are still unclear, modifying the test-bench operating protocol so that the signal generator provides a $4.4\,V_{pp}$ (V_{pp}: peak-to-peak voltage) with an offset voltage of $-3.15\,V$ (so still operating the modulator exclusively in reverse bias), led to a situation where one of our test devices yielded zero response in the dark and non-zero response under illumination. An oscilloscope tracing corresponding to this latter example is shown in Figure 5.5.

5.3.3 Interpretation

The interpretation of these sets of results can be split into two categories: a) interpretation at the time of measurement and b) interpretation with hindsight.

At the time of measurement it was considered that taking the precaution of measuring results at the second harmonic would reduce the contribution of noise and pick-up both to the values and to the random variations of both S_{dark} and S_{light} by a great amount. This effect was to be the result of the fact that the signal generator creates a very weak second harmonic of the pure sinusoid it provides. Therefore, detecting the modulation-induced contribution to the difference ΔS between S_{dark} and S_{light} would become a much more significant component, if not the dominant.

In hindsight, we believe that whilst it is entirely possible that the contribution of electro-optical modulation to ΔS does increase by use of the second harmonic[6] we can still not conclusively prove that

[6] Albeit at the expense of lower absolute S_x values overall and therefore more susceptibility to noise generated throughout

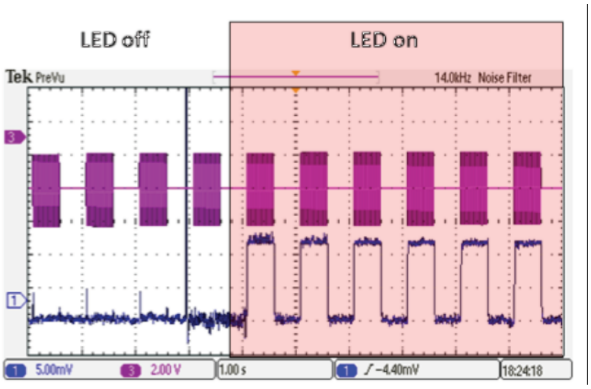

Figure 5.5: Typical oscilloscope tracing recorded during a proof-of-concept stage experiment. Reproduced from [2]. The device under test was device #4 from our C35 die #1 (NIN1-4), an N-well/sub type (NS) device. Experimental parameters: Input voltage offset: $-3.15\,V$. Peak-to-peak input voltage magnitude: $4\,V_{pp}$. Stimulus frequency: $1\,kHz$. Stimulus square wave envelope frequency: $1\,Hz$. Lock-in frequency: $2\,kHz$. Integration time: $3\,ms$.

ΔS is dominated by electro-optical phenomena. Specifically, we believe that a more accurate way of decomposing the contributions to S_x values would take the following form:

- S_{dark}: Pick-up from the effects of charge shuttling in and out of the pn-junction under test plus noise.
- S_{light}: Pick-up from the effects of charge shutting in and out of the pn-junction under test plus, electro-optical modulation, plus noise plus an illumination-induced photodetector small-signal impedance change that scales the response of the photodetector to pick-up (see section B.2.4 of the appendix for a more detailed explanation).

We notice first of all how pick-up is now no longer attributed solely to the signal generator, but is seen as a result of charge shuttling in and out of the test pn-junction. This is significant because the second harmonic of the signal that creates pick-up is now generated at the test device and will therefore will be considerably stronger than would be estimated by simply attributing it to the signal generator. However, this reinterpretation does not change the fact that the difference between S_x values in the dark and under illlumination are still present and can be attributed to electro-optical modulation.

Secondly, we notice how possible contributions to ΔS can now originate from a difference in how the photodetector responds to pick-up. In combination with the possibility that pick-up at the second harmonic may potentially be quite strong, we reach the conclusion that ΔS may be the result of either electro-optical modulation, or photodetector small-signal impedance change or a combination of both.

It was on the aftermath of such results that we decided to re-design our test bench so that it becomes more accurate and test conditions become more controllable. We then ran a new set of experiments with the aim of this time characterising the phenomenon rather than just observing its existence.

the system.

5.4 Test-bench design and operation

Measuring electro-optical modulation in CMOS integrated circuits required the development of a specialised test-bench. This section begins with a brief, conceptual overview of the said test-bench and then proceeds to elucidate the rationale and uncover all the technical details involved in building and operating it.

5.4.1 Brief conceptual overview

The test-bench employed was based on the transmittance mode configuration from chapter 4. Thus light was emitted from an outlet at the top-side of the modulator die, crossed the width of the die and hit a photodetector located contralaterally to the emitter outlet with respect to the die. Throughout all experiments the light emission outlet was kept in a rigidly fixed position whilst the photodetector was locked into a specified and immutable position with respect to the modulator. Therefore the modulator-photodetector assembly could be moved as a whole. Thus our test-bench featured the capability of aligning modulator devices to a common location where the irradiance received from the light source was consistent between tests whilst also keeping the photodetectors in a consistent position. Illustrations of the chip packaging method employed and the full test bench can be found in Figures 5.6 and 5.9 on pages 175 and 180 respectively.

Experimental protocols designed for the operation of the test-bench were also developed in order to ensure consistency of results and the extraction of meaningful information from the set-up. Notably, the free-carrier absorption phenomenon that we are trying to exploit is extremely weak and often overpowered by sources of noise. For that reason the strain placed upon experimental protocol design was not insignificant. Elaborate techniques, such as single-frequency signal processing (for example input: pure sinusoid - output: lock-in amplifier) had to be used in order to extract any useful data from the set-up.

5.4.2 Rationale behind modulator test bench design

A test bench needs to be designed that will be able to consistently detect the modulation of light that passes through our test devices. In theoretical terms the test bench must involve:

A) A stable light source: The light source performance in terms of transient noise and also the spectrum of the noise will play a large part in determining the minimum voltage variation that can be applied across the modulator while maintaining decent reliability in signal transmission. I.e it is a very important component in establishing the noise floor because it determines the quality of what we shall hence forth call the ‘reference beam’; the baseline beam against which deviations hopefully induced by modulation action alone can be measured. Moreover, the light source must suit the operating range of wavelengths that the modulator can reasonably process.

B) A pathway from light source to modulator: This can either be a simple air-gap in the case where the light source is positioned such that emitted light naturally reaches the target, or it can be waveguide-based, i.e. an optic fibre or fibre bundle. This would allow the flexibility of positioning the light source in a more convenient location and then simply guiding the emitted light to its target. However, other issues also come into play. The light beam that will exit the emitter assembly, either directly or via an

optic fibre, will likely be subjected to certain optical effects that distort its intensity profile. For example our reference beam may end up exiting an optic fibre at a 60^o fan-out angle with a beam diameter upon exit of $1\,cm$. Such beam profile is likely to lead to our light beam being largely underused when its task is to provide reference illumination to a 300×300 micron device. Thus beam profile control optics may need to be employed along the light source-to-modulator path.

C) A way of aligning the reference beam to the target modulator: This can be physically achieved via a positioner, manual or motor-controlled, that can move either the reference beam or the target relative to one another with great consistency. The aim is to ensure that the beam profile illuminating the active modulator area is consistent between all trials across all devices. However, simply ensuring that the modulator is entirely within the illuminated area of the reference beam will generally not be enough to guarantee this consistency unless the beam profile is uniform. It is for that reason that collimating the reference beam so that it acquires a more uniform beam profile throughout or taking care that the beam profile convoluted with a uniform averaging mask the size of the target modulator can be described by a distribution with only one local (which by inference will also be the global) maximum will alleviate the problem. The first option will do so by rendering a crude alignment, whereby the target device is simply be allowed to lie within the illuminated area of the reference beam, more tolerable as an alignment method. The second option will unlock the opportunity to align each individual device to the point of peak illumination on the alignment plane[7], an operation easily achieved by measuring the photocurrent generated by each modulator when used in power harvesting mode and then tracking the maximum manually. Note: either assumptions about the shape of the beam profile or measurements of the said beam profile need to be taken into account when interpreting data from devices that have different footprints.

D) A photodetector capable of capturing modulated light: Trivial though it may seem, the photodetector's characteristics, such as spectral sensitivity and dark current will affect overall system performance. Essentially the operating wavelength of the photodetector must match those of the light source and the modulator and the inherent noise of the device will play a role in the determination of the noise floor system-wide.

E) A pathway from modulator to photodetector: This can be done either by proper positioning of the photodetector relative to the modulator, or by coupling the light exiting the modulator into a waveguide. Care must be taken to collect at least all the light that has traveled through the modulator as that component carries all transmitted information (at least ideally). Any other, unmodulated components, however, can be safely ignored and allowed to dissipate.

F) Instrumentation that can control the light source, control the modulator, sense an input from the modulator (potentially for alignment purposes) and sense an input from the photodetector. These instruments need to add negligible noise to the system, if possible, as well as satisfy a range of other requirements related to their electronic characteristics and the test patterns used for assessing modulator performance (e.g. maximum operating frequency, current drive capability, dynamic range, available output waveform patterns etc.).

G) Test protocols: Aside from the physical aspect of the test bench there is also the operational aspect. Elements A-F need to be utilised in the best possible way in order to keep sources of error at bay and extract clean, meaningful results.

[7]The plane along which the micropositioner can shift the modulator die.

For convenience, we will split the analysis of the test bench into only four components: chip packaging and mounting, electro-optical set-up, instrumentation and test protocols. The packaging and mounting component will deal with the physical aspects of positioning the modulator in such way that it allows light to travel through it while itself being in a very rigid and well-controlled position. The electro-optical set-up section will consider the test-bench from the point of view of a light beam traveling from light source to detector and will thus touch upon all alignment issues. The instrumentation part will deal with all the interfacing to light sources, modulators and photodetectors. Finally, the test protocol section will explain the experimental procedures used to extract 'raw' data from our devices. Due to the unique challenges inherent in each of these components, they have evolved largely independently of each other.

Note: In this section we shall examine the various components of the final test bench only but mention key learnings from previous attempts as well. Finally, an overview of the final set-up will be given.

5.4.3 Device packaging and mounting

The packaging and mounting of each device revolved around a plastic package and a PCB respectively. The entire ensemble was designed under the idea that the reference light beam should be able to enter the die from the top side at an angle perpendicular to the surface of the die and exit from the bottom side at the same angle; an idea that warrants packaging the die on an optically transparent substrate (between ≈ 350 and $2000\,nm$ wavelengths). Similarly the PCB should either have a hole through which the light can pass or accommodate both the modulator and the photodetector.

Given these constraints the packaging solution chosen involved a standard, square plastic package with a $2.9\,mm$ hole drilled through its dead centre. A PLCC-type package with 64 pins served as the main packaging frame, a decision made on the basis of both bond-pad numbers and distribution across all designed test dies (Ninja, Teddy, Svejk and Bean) and with a view towards offering us the capability to quickly and effortlessly change dies on the same set-up. Once the $2.9\,mm$ hole had been drilled through the package, a square glass cover cut to a size not much larger than the minimum necessary to fully circumscribe the footprint of the largest die was glued over the hole with epoxy. The chip was then mounted on top of the glass substrate by use of a special, optically transparent epoxy (302-3M by Epo-Tek) since light would have to travel through it under the concept of the set-up. Subsequently, the assembly was wire-bonded.

Another important factor in choosing a suitable package was the presence of a cavity within which the chip could lie and the depth of that cavity. Many standard packages feature such cavities in order to accommodate a lid that can cover the cavity with the chip attached inside without at the same time disturbing the bond wires. The PLCC-64 package used did feature such cavity and it was of sufficient depth in order to allow the attachment of a glass lid over the entire extent of the cavity and safely cover all underlying components. Such glass lid was used so that the dies and their bond-wires would be protected against any mechanical damage and the overall assembly would be as shielded as possible against the ingress of dust particles and other contaminants. A conceptual diagram of the finalised packaging is shown in Figure 5.6.

The insertion of glass above and below the die will naturally introduce refractory surfaces ($Air - Glass, Glass - Air, Glass - Si$, although there will be some epoxy between the glass and the bottom of the Si die). However, since these refractory surfaces will be present in all packaged dies their effects in

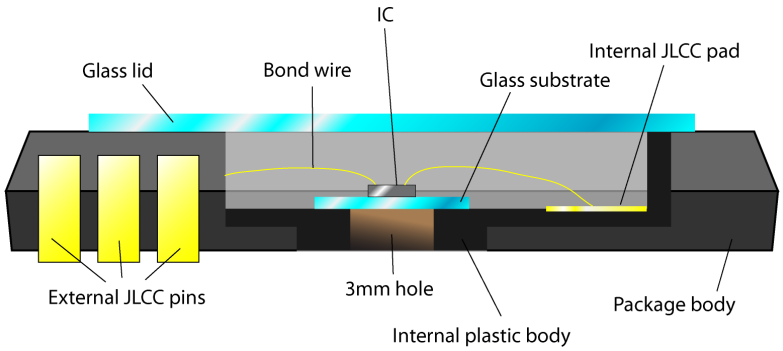

Figure 5.6: Packaging technique used to house all designed dies. The main body of the package has been exposed at its central location to show a cross-section of the interior. A glass lid encases the IC die into a sealed environment. The hole underneath the die is similarly covered by glass. Bond wires link the die to the internal pads that lead to the package's external pins.

each case are expected to be relatively similar both in terms of overall optical losses and in 'spectrum in' vs 'spectrum out' distortion[8]. As such in a relative, qualitative study of our electro-optical devices the effects of these refractive surfaces can be expected to play an insignificant role. Further study is required in order to definitively prove or disprove this statement.

In terms of mounting, a PCB-based set-up was used whereby pairs of identical PCBs, arranged directly one on top of the other constituted a single testing 'unit'. These PCBs were intended to be aligned one under the other so that their 'top views' would coincide perfectly. Two manifestations of this concept were used in practice: Set-up I, where photodetector and modulator reside on different PCBs and set-up II where they share a single PCB, the other one being relegated to the role of providing spatial separation between the solder-ridden bottom side of the top PCB and the motorised positioner platform. In set-up II the mechanical separation was achieved by means of rubber feet whilst set-up I used long plastic pillars. Set-up I was the last to be attempted.

In set-up I (the 'oil rig'), the bottom PCB was used to house the photodetector and an SMA output whilst the top PCB served as a base for the PLCC holder that would itself host modulator dies and corresponding BNC terminals, as well as switches or dual headers for routing various signals to their respective BNC terminals selectively. The pair of PCBs were held at a distance of approximately $90\,mm$ between them by four nylon pillars with metallic threaded housings either end in order to allow their secure attachment to the PCBs by metal screws. The footprints of the photodetector and and PLCC holder on the twin PCBs were placed in such way that the components they represent share the same geometric centre (manufacturing variation notwithstanding). Moreover, the tight attachment of the pillars to the PCBs via countersunk, metal screws ensured that the relative movement between PCBs was highly restricted. This created a situation whereby the modulator and the photodetector were latched in a convenient position that would allow light to cross the modulator at an angle perpendicular to its surface and after traveling a further $90\,mm$ hit the photodetector. The reason between the $90\,mm$ separation was to block most of the pick-up noise that would otherwise feed directly from the signal generator that supplies the test wave-form to the photodetector and the downstream amplifiers (see section on instrumentation).

[8]Unfortunately we have no readily available data on the exact absorption spectrum of the glass we used for our lids as the lids did not come with such documentation.

Despite all efforts to maintain proper alignment, the 90 cm separation between top and bottom PCBs and the nylon structure of the pillars did affect alignment to some extent. Thus, set-up I was the version that prioritised low interference at the expense of alignment. Determining misalignment by any reliable, quantitative standard was very difficult to achieve, but direct observation showed that light passing through the well-centred hole of an empty package did hit the photodetector roughly at its centre. Note: PCB designs were specific to each die, but other than that the mounting concept applied was the same throughout. PCB diagrams for each die can be found in the appendix.

In set-up II (the 'sandwich'), the top PCB carried both modulator and detector. The PLCC holder was specifically chosen so that once inserted, the modulator package would sit on an elevated plastic 'platform' with the photodetector comfortably fitting underneath the PLCC holder. A hole had to be drilled in the plastic platform so that the light crossing the modulator could proceed to hit the photodetector buried underneath. Both the through-holes punched into the PLCC holder and the photodetector were positioned to ensure a very tight fit that would force the overlying PLCC package and underlying photodetector to align -effectively- perfectly, save for PCB and photodetector/PLCC holder manufacturing variation. Set-up II thus emphasised strict alignment over interference from the signal generator. The closer proximity of the PCB tracks carrying signals to and from the modulator and the photodetector respectively did indeed lead to pick-up that was orders of magnitude above the corresponding values observed in set-up I.

Importantly, in both set-ups the die was tightly wedged into its PLCC-type carrier in order to make sure that the surface of the die remains parallel to the PCB surface. The fact that the PLCC carrier was designed to keep PLCCs such as the one that hosts our die means that the fit was always snug. This is important because any mis-orientation of the die surface would effectively affect the irradiance at each modulator device: each square micron of layout area would receive a reduced amount of optical power in accordance to:

$$I_\theta = I_0 cos(\theta) \tag{5.3}$$

where θ is the angle discrepancy, i.e. the quantity that shows by how much the surface of the die is not parallel to the surface of the host PCB[9] and I_x is the irradiance at angle discrepancy x in degrees.

Key learnings

Our attempts at constructing a viable packaging and mounting system showed that:

- Tightly-fitting custom-hole set-ups do not tend to work: This draws upon experience with direct mounting of the modulator die on the PCB. A notch was designed in the PCB within which the die was supposed to fit perfectly and in the process also align to the other components with a very high degree of accuracy. In reality, variation in the manufacturing of the PCBs made the notch unsuitable for the purpose. Instead, standard, round holes can be positioned with sufficient accuracy for a tight-fitting set-up.

[9] More rigorously defined as the angle between a vector coplanar to the host PCB and perpendicular to the line where the PCB and die surface planes meet and a corresponding vector coplanar with a corresponding vector, coplanar to the surface of the die.

- Pick-up noise can be extremely severe: Pick-up noise was a major issue throughout the entire project. It was established that pick-up would be very strong whenever tracks linking the photodetector to its SMA outlet would be routed in close proximity to tracks linking into modulator devices (set-up II).

5.4.4 Electro-optical set-up

The electro-optical part of the set-up can be explained by following the light trajectory from emitter to detector.

Light was generated in a Dolan-Jener MI-150 light source featuring a broadband, 150W 'EKE' halogen light bulb. The light-bulb emits light in almost the entire visible spectrum with significant components in the NIR range (see Figure 5.7). Once emitted, a 0.06 inch optic fibre bundle (Edmund Optics - 39-365) captures part of the light and carries it to a collimator. The construction of the light source means that the optic fibre is not preceded by any lens assembly that would allow more of the emitted light to be coupled into it. Instead much of the light intensity is wasted inside the illuminator, within the chamber where the light bulb is kept. The optic fibres allow for good optical transmission between the wavelengths of $0.4\,\mu m$ and $2.1\,\mu m$ with a dip around the $1.3 - 1.4\,\mu m$ range. This is expected to lead to slightly compromised system performance as the 'notch' range lies within the wavelengths that are useful for modulation. Upon exiting the fibre bundle, the light travels into the collimator (Edmund Optics - 58-836), where the outgoing beam divergence is reduced to 7.7^o. This is important in order to keep the beam profile as uniform as possible and avoid losing optical power. The last step before light is allowed to enter the modulator die consists of the filter that is positioned in front of the collimator. A $1300\,nm$ cut-off wavelength long-pass filter (ThorLabs - FEL1300) can be either introduced into or removed from the beam path. This is important because with the filter removed alignment can be performed visually (due to the presence of visible components in the beam), whilst with the filter present visible components of the beam are prevented from interfering with both modulator and photodetector (for example by modifying their small signal impedances).

Note: Because of poor beam shaping at the collimator level typically the entire $2.9\,mm$ diameter exposed surface of the photodetector was illuminated at all times. This means that only a very small fraction of the light reaching the photodetector is being modulated. As an example let us consider the case where a $0.5 \times 0.5mm$ device is active. In this case we have a modulator covering $0.25\,mm^2$ and a photodetector collecting photons from an area of $\approx 6.605\,mm^2$. In the case where we assume that the baseline transmittance over the entire photodetector collecting surface is similar this means that the baseline light intensity is artificially increased by approximately 26 times. Within the context of measured fractional modulation depth ($\frac{\Delta I}{I}$ where I is the baseline photocurrent and ΔI the difference in photocurrent between some chosen 'high' and 'low' bias conditions) this means a denominator inflated by a factor of 26 and therefore a modulation depth at any given 'high' and 'low' bias voltage conditions is reduced by the same amount. Lack of better available optomechanical components barred us from addressing this issue effectively.

The next obstacle encountered by the beam is the modulator die. The rigid assembly hosting modulator and photodetector was forced into alignment with the incoming light beam by means of an X-Y

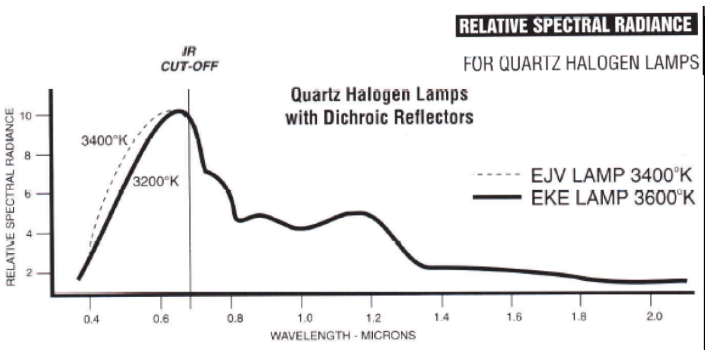

Figure 5.7: Spectrum of light source used for modulation experiments. Adapted from the datasheet as found on the website of the manufacturer that produced our MI-150 illuminator, at http://www.dolan-jenner.com/ PDFS/ EKE_EJV_SpectralCurve.pdf

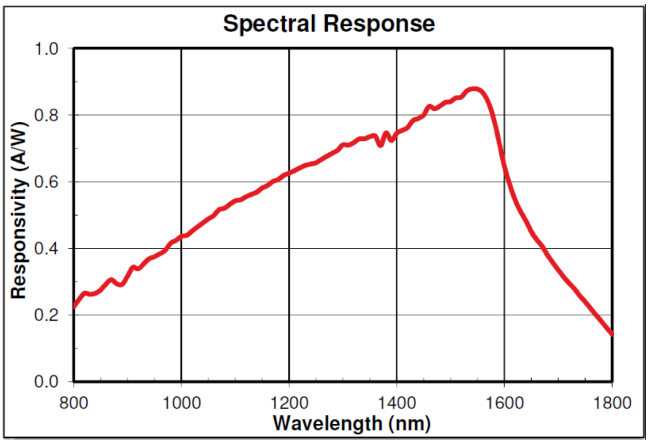

Figure 5.8: Sensitivity spectrum of the photodetector used for modulation experiments. Adapted from the datasheet as found on the manufacturer's website at http://www.thorlabs.de/ Thorcat/ 13800/ FDG03-SpecSheet.pdf

micropositioner (ThorLabs - MTS25/M-Z8) capable of incrementing/decrementing the X or Y coordinate of the target by approx. 100 microns with a repeatability in the range of 2 microns. Once alignment was achieved, as assumed by the discovery of a maximum in the photocurrent of the test modulator in power mode, the light could proceed through the die and hit a Germanium photodetector (ThorLabs - FDG03) with a sensitivity to light between approx. 800 and 1800 nm wavelength (see Figure 5.8 for full spectrum). The photodetector features a 3 mm diameter active area and a 1 μA dark current under 1 V reverse bias. Note: A diagram of the full system is shown in the next section, after the instrumentation set-up has also been described (Figure 5.9 on page 180).

Key learnings

Attempts to construct a suitable electro-optical platform resulted in the following observations:

- Simple, semiconductor lasers are not necessarily suitable reference light beam candidates: This stems from tests carried out with a temperature-stabilised, 1.5 mW solid state laser. When setting both a simple LED and our laser to an optical output of roughly 1.5 mW optical (as determined

by appropriate biasing according to the datasheets) the noise floor as measured by our set-up was in the region of $-110\,dBV$ for the LED and $-92.5\,dBV$ for the laser. This can prove fatal if trying to extract a weak, narrow-band signal. The noise floor in a similar set-up with the halogen source was $-116\,dBV$ but the overall power of the beam that hits the photodetector is unknown, partially due to beam divergence.

- No beam can be assumed to be collimated: At least not with cheap equipment. This stemmed from the fact that a range of collimators tested (fibre-coupled and fibre-bundle-coupled) all produced beams that resulted in non-uniform irradiance on a test surface. In all cases at least one local maximum was present, indicating that at least part of the variation in irradiance was not due errors in the positioning of the test photodetector between trials. The test photodetector itself was a $25 \times 25\,\mu m$ photodiode.

5.4.5 Instrumentation setup

The instrumentation used to interface with the set-up, introduce useful signals and extract meaningful information can be split along the lines of what part of the set-up it is feeding. Thus there are instruments that regulate the light source, feed or extract signals from the modulators and extract signals from the photodetector.

The light source is in itself an instrument, but the only control available for the regulation of the light output is an intensity regulator knob with ten different levels. For all experiments the light source was used at full power.

The modulators were connected in different configurations for alignment and modulator operation purposes. In alignment mode a source-meter (SMU) (Keithley - 2602A dual SMU) was connected to the modulator device and forced the voltage across the junction to $0\,V$. Then, the short-circuit photocurrent was measured as various locations were tested in search of a maximum. In modulator operation mode each device was connected to a signal generator (Agilent - 33522A) with the capability of producing sinusoidal inputs of given frequencies, amplitudes and offsets, amongst other waveforms. Both SMU and signal generator physically interfaced to a BNC termination on the PCB hosting the modulators.

The photodetector was connected to a low noise current amplifier (FEMTO - DLPCA-200) via a BNC cable. The amplifier was configured to maintain the voltage across the photodetector at a relatively constant bias voltage, close to $0\,V$ (when tested, the bias voltage across the photodetector was found to be in the single-digit mV figures). This allowed a current approximating the photodetector short-circuit photocurrent to enter the amplifier circuitry and be magnified by a factor of 10^4, a value which was empirically found to give a suitable range, not wastefully small and not large enough to saturate the circuitry down-stream. The output of the amplifier was split two-ways via a BNC T-junction, whereby one copy went directly to the oscilloscope (Agilent - DSO7014B) and the other was fed into a lock-in amplifier (Stanford instruments - SR810). The lock-in amplifier added the capability of extracting specific frequency components from the data being recorded at the photodetector site. The output of the lock-in amplifier was sent to the oscilloscope. The full test-bench system can be seen in Figure 5.9.

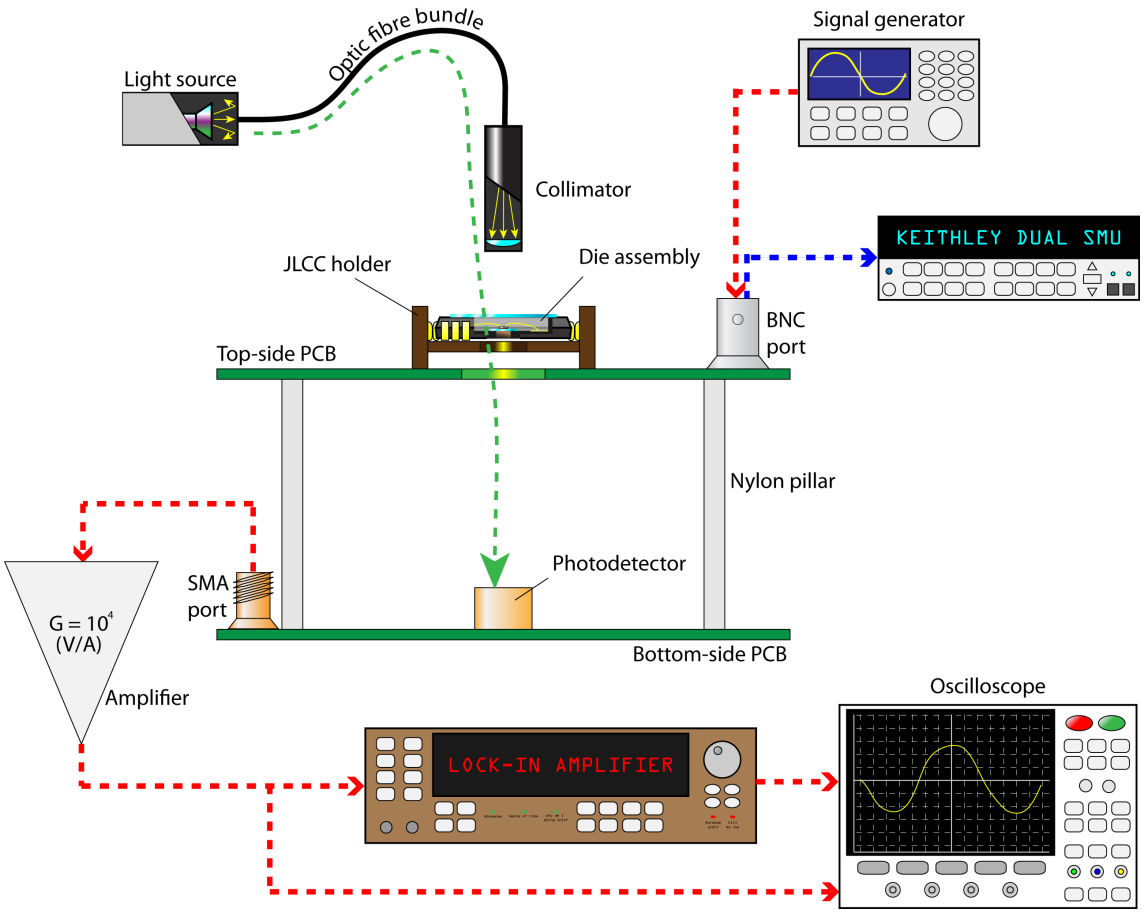

Figure 5.9: Overall test-bench set-up (set-up I variant). Various components mentioned throughout this section are labeled on the image. A more detailed view of the die assembly alone can be found in Figure 5.6. Red, dashed arrows show instrument connections during modulator operation whilst blue, dashed arrows show connections during the modulator alignment procedure. The green, dashed arrow indicates the path that the light takes during the procedure.

Key learnings

The development of a suitable instrumentation set-up has yielded the following learnings:

- Light sources need to achieve thermal equilibrium, or results will be subjected to drift: This holds true for at least the halogen light bulb we used in our final set-up, as well as for the LED- and laser-based systems employed previously. As such, the instrumentation controlling the light source can not be used as a reliable indicator of emitted light intensity. Instead, testing with a photodetector will yield a more accurate estimate assuming it will not be subject to significant thermal quenching itself due to heating from the light source.
- Signal generators may not always source what is requested of them (in terms of voltage): The signal generator must have the ability to drive the load it is supposed to be servicing and a good way of determining whether that is the case or not is to connect a copy of the output of the signal generator to an oscilloscope (via a T-junction for example).
- Theoretical values of voltage at biased nodes across any circuit may not always correspond to real values: This was discovered by measurements taken from the equivalent of different 'taps' along the BNC cable assembly connecting the photodetector to the current amplifier. For more information see B.2.4.

5.4.6 Test protocols

The test protocols are to the test-bench what software is to a computer: they have no physical presence, but play a key role in ensuring that the physical part of the set-up functions properly and yields meaningful results. The modular nature of the test-bench means that test-protocols too can be considered separately for light-source, modulator and photodetector.

Light source protocol: The light source was kept at full power and allowed to thermally stabilise for a few minutes before measurements were taken. It was found that the halogen light source was generally quick at settling to a stable optical power output once activated (generally < 15').

Modulator protocol: Each modulator device was subject to a strict and uniform testing procedure consisting of an alignment phase and a measurement phase. In the alignment phase the SMU was connected to the modulator, set to $0\,V$ bias and measured photocurrent. Small step changes in location were then implemented manually via the micro-positioner controllers. A maximum illumination point was eventually reached, which was assumed to be a global maximum if shifting the modulator at least two steps in any direction ($\approx 200\,\mu m$ absolute shift) resulted in a fall in intensity with respect to the obtained peak value. The measured photocurrent at the peak point was recorded and will be hence-forth referred to as the 'alignment photocurrent' (or 'alignment current' for brevity) of the device. When light-beam and PCB are aligned in such way that a certain device is capable of generating its peak photocurrent value, then we shall state that the said device is in 'full alignment'.

Once full alignment was confirmed the device was disconnected from the SMU and linked to the signal generator instead. Empirical tests indicated that an $8\,kHz$ signal would comfortably lie in a

frequency domain region characterised by relatively low noise and therefore an $8\,kHz$, $2\,V_{p-p}$, $\pm 1\,V_{offset}$ was applied to each modulator. The sign of the offset voltage was determined by the polarity of the pn-junction modulator, i.e. how its terminals were connected to the BNC core and cladding conductors. All devices were operated in the reverse bias range between 0-bias and $2\,V$ reverse bias; reasonably 'safe' values that were deemed unlikely to cause avalanche breakdown or damage the test structures in any way. Note: operating the devices in reverse-bias still implies that a certain amount of reverse bias current (and if part of the illumination spectrum is within the region where inter-band absorption becomes significant also photocurrent) is shuttling through the depletion region creating leakage currents. The presence of these leakage currents means that the depletion region is not entirely optically transparent for NIR wavelengths, but since such currents are extremely small we decided to ignore the phenomenon for the purposes of this study. Obtaining hard evidence on whether such currents do affect modulation depth (absolute and fractional) to a significant degree can be the target of future work.

Arguably devices consisting of different junction types, of different sizes, featuring different junction capacitances will show different impedances at various levels of voltage bias and frequency, but an important part of this project was to determine how different devices react to similar input signals in order to determine which would be the best performer given similar signal constraints (in terms of headroom etc.). The signal generator was allowed to provide the test signal for as long as was necessary (see photodetecor protocol), after which the device was disconnected from the signal generator altogether.

Photodetector protocol: The photodetector was fed into a trans-impedance amplifier that was empirically found to re-map the input signal range to the best output signal range when set to a gain of 10^4. Thence forth two copies of the signal were generated and routed to different instruments.

One copy of the signal was then fed directly into the oscilloscope where a standard transient trace was displayed along with its fast Fourier transform (FFT). This was intended to indicate a peak at $8\,kHz$ appearing as a result of modulation. The observed appearance of a large peak at exactly $8\,kHz$ after amplification was an indication that any non-linearities introduced by our amplifier did not dramatically affect the spectral content of the signal arriving from the photodetector. This, coupled with the fact that we are simply trying to detect the presence or absence of modulation (digital signal transmission), means that any non-linearities in the amplifier are only expected to significantly affect signal quality when the modulation on/off difference is close to the limit of detection. In such situation any non-linearities introduced by the amplifier may make the difference between a measurable and a not measurable signal.

The second copy of the signal was fed into the lock-in amplifier. As is known, the lock-in amplifier can lock onto the $8\,kHz$ component and yield a constant value that depends on the: a) amplitudes of the incoming and reference signals. b) relative phase of the incoming and reference signals and c) Period of integration. Analytically the output of a lock-in amplifier is given by:

$$V_{out}(t) = \int_{t-T}^{t} f(t) A_{ref} \sin(\omega_{ref} t + \phi_{ref}) dt \tag{5.4}$$

where $V_{out}(t)$ is the output voltage as a function of time, T is the integration period, $f(t)$ the input signal waveform, A_{ref} the amplitude of the internal, reference signal, ω_{ref} the lock-in frequency (angular), ϕ_{ref} the phase of the reference signal (as determined with reference to some fixed time frame t_0) and t represents time.

Taken for each individual frequency component of $f(t)$ equation 5.4 simplifies to:

$$V_{out} = \int_{t-T}^{t} A_{sig} \sin(\omega_{sig} t + \phi_{sig}) A_{ref} \sin(\omega_{ref} t + \phi_{ref}) dt \tag{5.5}$$

where the equation is the same as 5.4, but $f(t)$ has been now replaced by a function of exactly the same form as the reference signal. By merging the amplitude constants A_{ref} and A_{sig} and careful positioning of the time reference point t_0 equation 5.5 further simplifies to:

$$V_{out} = K \int_{t-T}^{t} \sin(\omega_{sig} t + \phi) \sin(\omega_{ref} t) dt \tag{5.6}$$

where the constant K eliminates the absolute amplitude of the reference signal and the time frame has been positioned to null the phase of the reference signal. If we set $\omega_{ref} = \omega_{sig}$, equation 5.6 clearly demonstrates the dependencies of V_{out} on absolute and relative amplitudes of incoming and reference signals (K), their relative phases (ϕ) and the period of integration (T).

This creates the problem that if the lock-in amplifier is set to sensing exactly $8\,kHz$, then it may lock on to the incoming signal at any phase, meaning that the output will be a function of both relative phase and incoming signal amplitude (note: the reference signal amplitude and integration period are kept constant throughout). Manually adjusting the phase of the reference signal in search of a maximum can yield an output that is a reliable function of incoming signal amplitude only, but there is a practically more effective way of determining that value.

Suppose we lock-in to a frequency very slightly shifted compared to the incoming signal frequency. Also, for simplicity let us assume that the incoming signal is for all intents and purposes a single sinusoidal input and that they start oscillating concurrently so that their relative phase shift is 0. This does not lead to loss of generality. Equation 5.6 thus becomes:

$$V_{out} = K \int_{t-T}^{t} \sin(\omega_{sig} t) \sin(\omega_{ref} t) dt \tag{5.7}$$

We now express the signal frequency as a sum between the reference frequency and a small constant:

$$\omega_{sig} = \omega_{ref} + \alpha \tag{5.8}$$

where α is the difference between reference and signal frequencies.

Equation 5.7 now becomes:

$$V_{out} = K \int_{t-T}^{t} \sin(\omega_{ref} t + \alpha t) \sin(\omega_{ref} t) dt \tag{5.9}$$

Using the trigonometric identity for sum of angles we obtain:

$$V_{out} = K \int_{t-T}^{t} \sin^2(\omega_{ref} t) \cos(\alpha t) dt + K \int_{t-T}^{t} \cos(\omega_{ref} t) \sin(\omega_{ref} t) \sin(\alpha t) dt \tag{5.10}$$

If we assume that the integration period is long enough and alpha is small enough to be considered constant throughout a period equivalent to T, then the integral of $\cos(\omega_{ref}t)\sin(\omega_{ref}t)$ over an integration period equaling T will tend to zero and can be thus eliminated. On the other hand, given the same assumption for the size of α, the first term of the integral dominates the equation, $\cos(\alpha t)$ becomes a a parameter that slowly varies with time and thus equation 5.10 simplifies to:

$$V_{out} = K\int_{t-T}^{t} \sin^2(\omega_{ref}t)sin(\alpha t)dt \approx Ksin(\alpha t)\int_{t-T}^{t} \sin^2(\omega_{ref}t)dt \tag{5.11}$$

which is simply the result of an ideal, 0-phase offset lock-in result multiplied by $\sin(\alpha t)$.

Indeed with an $8\,kHz$ incoming signal frequency, an α of $-1\,Hz$, incoming signal period of $125\,\mu s$ and integration period of $3\,ms$ our assumptions seem to hold true. $\sin(\alpha t)$ has a period of $1\,s$ and thus changes little over $3\,ms$, and the integration period is exactly 24 times the incoming signal period, which in itself should be enough to null the second term of the integral that we considered negligible.

In practical terms this means that measuring the peak-to-peak amplitude of the waveform coming out of the lock-in amplifier we have a direct measurement of K, which itself reveals qualitatively the strength of the $8\,kHz$ component as measured by the photodetector. In order to reduce uncertainties, 50 cycles of the lock-in output waveform were measured and their peak-to-peak (p-p) amplitudes were averaged. These averaged peak-to-peak amplitudes are single, scalar values that represent the outcome of any single test run and will be referred to as the 'A_{p-p} value' of the test hence forth. The peak-to-peak values for any individual cycle out of the 50 recorded in order to determine A_{p-p} will be referred to as singleton peak-to-peak values and marked by A_S.

Finally, to aggregate these results, tests were conducted where the amplitude of the $8\,kHz$ component was compared between conditions of no illumination and full illumination. More information about the validity of this technique and the usefulness of these results can be found in the section on result processing methodology and meaning (located within section B.1) and beyond.

Note: it is only natural that noise from any source in the vicinity of the $8\,kHz$ mark will interfere with our set-up and alter A_{p-p} values, but interestingly, because A_{p-p} values represent measures of peak-to-peak excursions in the lock-in amplifier output signal, any stray noise at exactly the lock-in frequency will be eliminated.

Note: Any distortion in the signal, be it from the amplifier, non-linearities in the photodetector or anywhere else will transform the 'ideal', pure $8\,kHz$ signal arriving from the signal generator (ideally) into a signal with a broader peak whilst still maintaining the primacy of the dominant peak (assuming no heterodyne mixing phenomena). Such distortion affects that narrow signal band that can influence the output of the lock-in amplifier. In practice it was found that a) the dominant peak remains at $8\,kHz$ after passing through the trans-impedance amplifier and that lock-in at exactly $8\,kHz$ does produce a quasi-DC signal (very slow drift corresponding to π phase shift over the course of minutes - unpublished observations) and b) locking to frequencies farther than $1\,Hz$ away from the nominal $8\,kHz$ value led to dramatically reduced lock-in amplified output. Both of these observations would indicate that any distortion introduced upstream of the lock-in amplifier is either small or cancels out somewhere along the chain. We therefore felt confident that locking-in at $1\,Hz$ off the nominal $8\,kHz$ peak would yield behaviour reasonably well-described by the equations seen in this section. Finally, whatever distortion was present may have affected all measurements similarly although this should be the subject of further

study.

Note: If the amplifier has a gain that depends on the amplitude of the input signal or the DC current level exiting the photodetector (which in turn depends on the amount of light falling upon the detector)[10] then we get different A_{p-p} values in the dark and under illumination even if we somehow accept that we have managed to block modulation (e.g. if we impose equivalent illumination upon the photodetector bypassing the modulator die). We decided to combat both issues by using an industrial-grade amplifier, however, there is no guarantee that such effects haven't played a role in determining the A_{p-p} values we obtained in the end.

5.4.7 Test-bench summary

From a practical perspective, the test bench was designed to provide stable illumination, maintain tight control over the relative positions of the light exit point (the collimator) and the modulator/photodetector assembly and keep modulators and photodetector in a state of reliable alignment. To this end a specialised chip packaging technique had to be developed as well as a way of forcing PCB pairs to stay mutually aligned yet at a comfortable distance in order to avoid interference. Standard instrumentation was sufficient for providing useful test signals (single-frequency AC) and recording important information concerning both the extraction of results and diagnostic tests to ascertain the functionality of the test-rig. The test protocols used ensured that under the condition that modulated light signal exists and is of sufficient strength, it could be measurable both by directly assessing the strength of its component in the FFT of the amplified photodetector signal and by locking to it via a lock-in amplifier. The latter method would essentially trade timing resolution for detectability of a weaker signal and substantial noise rejection.

5.5 Result interpretation

In this section, key results are extracted from lower grade data relegated to the appendix (B) and their implications are examined. The section is split into three main parts: a) general observations that affect the subsequent interpretation of results, b) relational results where comparisons between devices with different attributes are made and c) absolute results where the performance of our modulators is put into the context of an information transmission system.

5.5.1 General observations

Fundamental nomenclature

Throughout this chapter we will be making heavy use of modulator performance metrics that we have defined ourselves. These are:

- A_{p-p}: A measure of electrical signal intensity arriving from the photodetector within a specified, narrow frequency band corresponding to the input signal (see section 5.4.6 for formal definition).

[10] In plain terms if the small signal gain is different than the large signal gain in the regions of interest.

The characteristics of frequency band are irrelevant for the interpretation of results and common to all A_{p-p} values in this thesis. A_{p-p} values can be separately defined for experiments carried out under illumination or in the dark. note: A_{p-p} is a directly measured variable, unlike the B_{p-p} and C_{p-p} variables that follow, which are derived from A_{p-p}.

- B_{p-p}: The difference between A_{p-p} values obtained for a given device under illumination and in the dark, for otherwise exactly the same experimental conditions. Under ideal conditions, this would be a direct measure of electro-optical modulation (see note below).

- C_{p-p}: The difference between obtained and predicted B_{p-p} values in a special set of experiments carried out in section 5.6.2. The metric is listed here for completeness, but is fully defined and discussed in the relevant section.

Important note: A_{p-p} in the dark is thought to provide a measure of noise, pick-up and other parasitic activity whilst the corresponding value under illumination would include all of the aforementioned, plus the effects of electro-optical modulation. For this reason we can refer to A_{p-p} in the dark as 'background activity' or 'pick-up noise estimate' and A_{p-p} under illumination as 'aggregate activity'. Hence forth we shall assume that A_{p-p} is dominated by the pick-up noise (something we have confirmed by observing that when the signal generator stands quiet, A_{p-p} is nearly zero) and therefore will often refer to A_{p-p} under its 'pick-up noise estimate' moniker. As a direct consequence subtracting A_{p-p} in the dark from A_{p-p} under illumination (a quantity we have defined as B_{p-p}) should now give an estimate of electro-optical modulation. The specific conditions under which this would hold true are exposed farther down this section (see section on 'key assumptions' on page 187).

Origins and effects of pick-up on measurement procedure

Perhaps the most important observation that needs to be made concerns the level of baseline activity (pick-up strength) found in all devices and seen in their raw results. Closer examination of devices with unusually high baseline activity (notably dev. #5 from die Teddy #6, n-well/substrate junction i.e. 'TED6-5-NS') showed that the level of pick-up is strongly dependent on the layout of the PCB that was developed to match each die design. 'TED6-5-NS', notably has its signal routed via a track that passes very closely by the photodetector and that might be causing strong pick-up to contaminate the photodetector channel by adding a voltage perturbation to the photodetector. Whether such perturbations can be considered small signal or not is not entirely known. However, given that the signals leaving the photodetector have to pass through a gain 10^4 amplifier and will still not cause significantly large transients in the raw (not locked-in) signal[11] it is fairly safe to assume that the voltage variations imposed upon the photodetector are of a sufficiently small magnitude (mV and below - certainly much below the thermal voltage v_T) in order to be considered small-signal.

Crucially, it is these minute perturbations that combine with the small-signal impedance of the photodetector in order to create an illumination-dependent parasitic pick-up current that feeds into the preamplifier (see subsection B.2.4 for more details); the dominant contributor to all A_{p-p} values. This

[11] Specifically, the raw transient signal coming from 'TED6-5-NS' shows an overall oscillation magnitude of less than $200\,\mu V$ even though it has been amplified by a factor of 10^4.

is important because it means that not only pick-up itself, but also everything weaker, including electro-optical effects sited at the modulator die, can all be safely considered as manipulating the photodetector in the small-signal domain.

Key assumptions

Crucially, if B_{p-p} is to be considered a true representation of electro-optical modulation, then certain assumptions need to hold. Specifically, we have to assume that once we switch the light source on, the only additional signal that passes through the lock-in amplifier is the $8\,kHz$ component of the photocurrent caused by electro-optical modulation. Implicitly this demands the following:

- The A_{p-p} (dark) value remains constant or effectively constant from trial to trial. This is necessary because B_{p-p} is the difference between A_{p-p} under illumination and in the dark. If the assumption doesn't hold, then B_{p-p} also includes a component related to A_{p-p} (dark) drift.
- The only significant phenomenon affecting the output of the lock-in amplifier that appears upon switching the light source on is electro-optical modulation. If other light-dependent phenomena occur and are not at least dominated by electro-optical modulation, then they need to be quantified and removed somehow, or B_{p-p} will no longer be an accurate representation of electro-optical modulation.
- The presence of electro-optical modulation does not affect the underlying value of A_{p-p}, i.e. observed pick-up amplitude doesn't depend on illumination. This is equivalent to saying that optical effects and pick-up are independent phenomena whose effects on the lock-in amplifier output add according to superposition. The point made here is a very important special case of the point above.

Throughout this section we shall discuss results on the hypothetical basis that these assumptions hold true. It will be in section 5.6 that we examine these assumptions more closely, attempt to explain the phenomena that challenge them and describe what steps have been taken in order to minimise their effects. At the same time we will also attempt to estimate how the abolition of these assumptions might provide alternative explanations for the obtained results.

Sign of obtained junction coefficients

A brief, but interesting point needs to be made about the sign of our obtained junction coefficients. Since all coefficients are computed on the basis of B_{p-p}, i.e. the difference between A_{p-p} in the dark and under illumination, the sign of B_{p-p} determines whether the device modulation performance is 'positive' or 'negative'. In the case of a positive B_{p-p} electro-optical modulation is working synergistically with pick-up whilst in the case of negative B_{p-p} they work antagonistically to each other.

Although we have tried to understand what causes pick-up and what the exact nature of the relationship between pick-up and output is, we have failed to find consistent results. As evidenced by tables B.5, B.7, B.9, B.13 and B.15 in the appendix, only dies of type Svejk (H18) showed any real

consistency in the signs with only one outlier and all other devices reading negative B_{p-p} values. Based on this result we would speculate that under 'normal' circumstances electro-optical modulation should act antagonistically to pick-up and thus electro-optical modulation should be characterised by negative coefficients. Further study, however, is required in order to fully understand the nature of pick-up and its interaction with electro-optical modulation.

5.5.2 Relative comparison between devices

In this subsection we discuss relational data between devices. We examine the available data in order to see whether we can obtain any discernible information about the effects of technology, device type, geometry and inter-die variation on modulator performance. Given theoretical predictions we can expect that devices in smaller feature size technologies containing diffusion or long side-wall well-based junctions are likely to be the top performers. Moreover, given the large size of each device (footprint size) we can expect relatively small effects from random process variation, although any wafer-scale, low spatial frequency persistent variation in semiconductor properties will show.

In the sections below we attempt to present our high-level results in an order where previous results help understand subsequent ones. Thus we begin with the effects of inter-die homologous device pair variation on performance in order to gain some understanding of how not only inter-die variations, but also other extraneous effects, affect modulator performance from a trial of a test device to a trial of a homologous device residing on a different die belonging to the same design family. We then proceed to see how our obtained results help us link modulation performance to geometry with the aim of extracting a relation that describes the contributions of various junction components (specifically areal and side-wall junction components) to the resulting, observed modulation. We except to obtain coefficient values that link, for instance, estimated electro-optical modulation (B_{p-p}) to each micron of net side-wall junction perimeter or to each micron squared of the areal junction component. Next, we compare these obtained coefficients between different device types (NS vs pN areal components for example) before proceeding to examine differences between technologies.

Each of the following sections is organised in such way that as the reader progresses through it, the results become more and more compressed and distilled. At the end of each section a few tables and figures attempt to convey the core message of each section and the key learnings, leaving lesser details aside.

5.5.2.1 Effect of inter-die variation on performance

We were able to extract information from only five dies overall. Two covering the UMC13 technology, two covering IBM18 and one covering AMS35. Given more time that number was planned to be much larger and cover every technological node used multiple times. Thus, there is no data on inter-die modulation for the Ninja (NIN) design family, but corresponding data from Teddy die #5 (TED5) vs Teddy die #6 (TED6) and Svejk die #1 (SVJ1) vs Svejk die #2 (SVJ2) has been extracted.

The Svejks:

Tables 5.4 and 5.6 summarise the differences between Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) in the dark and under illumination respectively.

Table 5.4: Estimated pick-up noise levels (a.k.a. A_{p-p} in the dark) for Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Ideally no signal should be received in the dark. Differences in signal reception levels between dies are also shown in absolute (ΔA_{p-p}) and fractional form (F_{Ap-p}) according to the formula $F_{Ap-p}(a,b) = 2\frac{a-b}{a+b}$. The average fractional difference between homologous devices and the standard deviation between fractional differences are shown at the bottom of the table. The horizontal line breaking the table in two large sections separates devices with the large, 300 × 300 micron footprint (upper section) from those with the small, 200 × 200 footprint (lower section). Device # 12, N-well on substrate (SVJ-12-NS) has been removed as it was a clear outlier (reason unknown).

DARK					
		SVJ1	**SVJ2**		
DEV	**Type**	A_{p-p}	A_{p-p}	ΔA_{p-p}	F_{Ap-p}
		mV	mV	mV	%
SVJ-1	nS	1351.1	1273.4	77.7	5.92
SVJ-2	nS	1884.3	1855.1	29.2	1.56
SVJ-3-NS	NS	410.3	320.7	89.6	24.51
SVJ-3-pN	pN	450.8	404.1	46.7	10.93
SVJ-3-pNS	pNS	251.6	232	19.6	8.11
SVJ-4-NP	NS	413.3	316	97.3	26.68
SVJ-4-pN	pN	451.2	399.4	51.8	12.18
SVJ-4-pNS	pNS	270.4	238.9	31.5	12.37
SVJ-5	nS	263.5	221.7	41.8	17.23
SVJ-6	nS	292.7	231.3	61.4	23.44
SVJ-7	NS	1024.3	856.1	168.2	17.89
SVJ-8	NS	382.9	183.5	199.4	70.41
SVJ-9	NS	245.1	185.2	59.9	27.84
SVJ-10	NS	283.3	213.8	69.5	27.96
SVJ-11	NS	254.3	194	60.3	26.9
SVJ-12-3N	3N	752.2	623	129.2	18.79
SVJ-12-3NS	3NS	320.6	179.1	141.5	56.63
Average					22.90
σ					16.98

Dark: We will begin the analysis with data pertaining to measured A_{p-p} in the dark. From table 5.4 we notice that all Svejk #1 (SVJ1) devices show greater A_{p-p} values than their counterparts in Svejk #2 (SVJ2) with an average fractional difference (as defined before) of 22.90 %. Nevertheless the variation between these fractional differences is described by a σ value of 16.98 %, assuming the distribution of fractional differences is Gaussian. This assumption, however, does not hold[12], largely due to the different types of junctions involved. Still, it is useful to extract the mean for comparison with more detailed, junction type-specific data. This occurs because the standard deviation value bestows a sense of 'overall scatter' amongst data points and though a flawed and highly misleading metric in itself, it can be used for comparisons between experiments ran with the exact same sets of devices under different illumination conditions.

If we assume that a) charge shuttling in and out of each test device as a result of external biasing (from the signal generator) does play a role in shaping pick-up sensed at the photodetector (as opposed to the assumption that pick-up arises only from metal interconnects acting as antennae) and that b) different junction types are affected differently by inter-die variation, an assumption we have no reason to hold for untrue, then we can expect to see that different junction types will exhibit different average fractional A_{p-p} differences between dies. If assumption 'a' doesn't hold, then pick-up should not depend on any structure on the modulator die. If assumption 'b' doesn't hold, then any variation in dark A_{p-p} due to inter-die physical differences should be on average roughly the same for all junction types. Unfortunately there is not enough data to provide reliable statistical information about the behaviour of different junction types (notably standard deviation information), but we can group the Svejk (SVJ) devices by junction type in order to calculate their average fractional differences.

Table 5.5: Average fractional difference between estimated pick-up noise levels (a.k.a. A_{p-p} in the dark), grouped by junction type ($\overline{F_{Ap-p}}$). The last column is the number of junctions of each specific type taken into account. The fractional difference between two values is defined as $F_x = F(a,b) = 2\frac{a-b}{a+b}$ where 'x' sits in-lieu of (a,b).

Type	$\overline{F_{Ap-p}}$ **(%)**	**n**
	%	
nS	12.04	4
pN	11.55	2
3N	18.79	1
NS	31.74	7

Table 5.5 shows average fractional A_{p-p} differences by junction type. Data seems to indicate that NS type junctions lead to significantly more intense changes in A_{p-p} (dark) between dies than any other junction type. Moreover, diffusion-based junctions (nS and pN) show similar average fractional differences. Nevertheless, because of the low numbers of junctions tested for all junction types, these statements cannot be claimed with certainty. This is most noticeable in the fact that the Svejk design, dev. #1 (SVJ-1) and Svejk design, dev. #2 (SVJ-2) n-diffusion to substrate type (nS-type) pairs show very different inter-die A_{p-p} fractional differences than their Svejk design ,dev. #5 (SVJ-5) and Svejk design, dev. #6 (SVJ-6) homotypes[13].

[12] Relevant histograms not shown.

[13] Homotype devices: a set of devices that have the same junction structure (n-diffusion to substrate -nS-, N-well to substrate -NS-, p-diffusion to N-well to substrate -pNS- etc.). Homotype junctions: junctions sharing the same structure.

Illuminated minus dark: The same procedure can be applied to our extracted B_{p-p} values in order to obtain absolute and fractional differences between homologous devices. The corresponding table is shown as table 5.6.

Table 5.6: Estimated levels of electro-optical modulation (a.k.a. B_{p-p}) for Svejk #1 (SVJ1) and Svejk #2 (SVJ2) along with the differences between them. Some statistical metrics are also provided at the bottom of the table, such as the average fractional difference between homologous devices and the standard deviation between fractional differences. The horizontal line breaking the table in two large sections separates devices with the large, 300×300 micron footprint from those with the small, 200×200 footprint. The fractional difference between two numbers is defined as $F_x = F(a,b) = 2\frac{a-b}{a+b}$ where 'x' sits in-lieu of (a,b). Device # 12, N-well on substrate (SVJ-12-NS) has been removed as it was a clear outlier (reason unknown).

ILLUMINATED - DARK					
		SVJ1	**SVJ2**		
DEV	**Type**	B_{p-p}	B_{p-p}	ΔB_{p-p}	F_{Bp-p}
		mV	mV	mV	%
SVJ-1	nS	-99.1	-87.6	-11.5	12.32
SVJ-2	nS	-105.1	-91	-14.1	14.38
SVJ-3-NS	NS	-139.6	-62.3	-77.3	76.57
SVJ-3-pN	pN	-281.3	-207.3	-74.0	30.29
SVJ-3-pNS	pNS	-94.4	-30.5	-63.9	102.32
SVJ-4-NS	NS	-142.2	-61.4	-80.8	79.37
SVJ-4-pN	pN	-288.2	-215.6	-72.6	28.82
SVJ-4-pNS	pNS	-92.8	-31.9	-60.9	97.67
SVJ-5	nS	-93.3	-59.9	-33.4	43.6
SVJ-6	nS	-110.6	-64.6	-46.0	52.51
SVJ-7	NS	-76.9	-39.6	-37.3	64.03
SVJ-8	NS	-103.9	-63.6	-40.3	48.12
SVJ-9	NS	-96.0	-46.3	-49.7	69.85
SVJ-10	NS	-81.7	-42.7	-39.0	62.7
SVJ-11	NS	-86.1	-42.6	-43.5	67.6
SVJ-12-3N	3N	-58.6	-41.3	-17.3	34.63
SVJ-12-3NS	3NS	-60.4	-29.2	-31.2	69.64
Average					56.14
σ					25.73

It becomes immediately visible that the average absolute interdie variation grows significantly (almost doubles) whilst the variation of this average also grows significantly (to 25.73 %).

In table 5.7 we can see the results from table 5.6 grouped by junction type. Notably, the obtained figures are qualitatively similar to corresponding results in dark conditions. NS type junctions seem to be the most dissimilar between different dies in the pair with all other junction types showing much lower average fractional differences. Furthermore, diffusion-type junctions once again show very similar results although it must be noted that n-diffusion to substrate (nS) type devices Svejk design, dev. #1 (SVJ-1) and Svejk design ,dev. #2 (SVJ-2) once again behave very differently to their homotype Svejk design, dev. #5 (SVJ-5) and Svejk design, dev. #6 (SVJ-6) peers.

The Teddies:

Tables 5.8 and 5.10 summarise the differences between Teddy die #5 (TED5) and Teddy die #6

Table 5.7: Average fractional difference between estimated electro-optical modulation values (a.k.a. B_{p-p}) grouped by junction type ($\overline{F_{Bp-p}}$). The last column is the number of junctions of each specific type taken into account. The fractional difference between two numbers is defined as $F_x = F(a,b) = 2\frac{a-b}{a+b}$ where 'x' sits in-lieu of (a,b).

Type	$\overline{F_{Bp-p}}$	**n**
	%	
nS	30.7	4
pN	29.56	2
3N	34.63	1
NS	66.89	7

(TED6) in the dark and under illumination respectively.

Table 5.8: Estimated pick-up noise levels (a.k.a. A_{p-p} in the dark) for Teddy die #5 (TED5) and Teddy die #6 (TED6) along with the differences between them. Some statistical metrics are also provided at the bottom of the table, such as the average fractional difference between homologous devices and the standard deviation between fractional differences. The fractional difference between two numbers is defined as $F_x = F(a,b) = 2\frac{a-b}{a+b}$ where 'x' sits in-lieu of (a,b).

DARK					
		TED5	**TED6**		
DEV	**Type**	A_{p-p}	A_{p-p}	ΔA_{p-p}	F_{Ap-p}
		mV	mV	mV	%
TED-1	NS	267.7	344.5	-76.8	-25.09
TED-2	nS	239.4	327.7	-88.3	-31.14
TED-3	nS	264.5	347.0	-82.5	-26.98
TED-4	NS	295.6	359.6	-64.0	-19.54
TED-7-n3	n3	118.5	191.6	-73.1	-47.15
TED-7-3N	3N	347.1	344.7	2.4	0.69
TED-7-NS	NS	472.6	519.1	-46.5	-9.38
TED-8-n3	n3	167.4	175.3	-7.9	-4.61
TED-8-3N	3N	359.3	349.8	9.5	2.68
TED-8-NS	NS	801.6	795.0	6.6	0.83
TED-5-NS	NS	2726	2719.8	6.2	0.23
TED-5-3N	3N	621.4	608.7	12.7	2.06
TED-9-NS	NS	507.8	493.4	14.4	2.88
TED-9-3N	3N	282.9	262.8	20.1	7.37
Average					-10.51
σ					15.97

Dark: The Teddy die #5 and #6 (TED5-TED6) pair shows a $-10.51\,\%$ average fractional discrepancy between homologous device pairs with a spread of $15.97\,\%$. The maximum discrepancy was observed for Teddy design, device no7, n-diffusion on triple well junction (TED-7-n3) with a value of $47.15\,\%$.

Grouping results by junction type yields table 5.9. Different junction types show different F_{Ap-p} values but interestingly diffusion-based junctions show relatively similar results (n-diffusion on substrate (nS) and n-diffusion on triple well (n3) types).

Illuminated minus dark: The average fractional difference between homologous dies grows significantly

Table 5.9: Average fractional difference between estimated pick-up noise levels (a.k.a. A_{p-p} in the dark) grouped by junction type ($\overline{F_{Ap-p}}$). The last column is the number of junctions of each specific type taken into account. The fractional difference between two numbers is defined as $F_x = F(a,b) = 2\frac{a-b}{a+b}$ where 'x' sits in-lieu of (a,b).

Type	$\overline{F_{Ap-p}}$	**n**
	%	
nS	-29.06	2
n3	-25.88	2
3N	3.2	4
NS	-8.35	6

compared to the dark case and achieves a value of $-22.76\,\%$. The spread in fractional variations, however, takes the unexpectedly high value of $135.77\,\%$, which hints towards the fact that inter-die inconsistencies affect modulator performance very heavily in this technology.

Table 5.10: Estimated electro-optical modulation levels (a.k.a. B_{p-p}) for Teddy die #5 (TED5) and teddy die #6 (TED6) along with the differences between them. Some statistical metrics are also provided at the bottom of the table, such as the average fractional difference between homologous devices and the standard deviation between fractional differences. The fractional difference between two numbers is defined as $F_x = F(a,b) = 2\frac{a-b}{a+b}$ where 'x' sits in-lieu of (a,b).

ILLUMINATED - DARK					
		TED5	**TED6**		
DEV	**Type**	B_{p-p}	B_{p-p}	ΔB_{p-p}	F_{Bp-p}
		mV	mV	mV	%
TED-1	NS	12.4	30.0	-17.6	-83.02
TED-2	nS	5.1	36.6	-31.5	-151.08
TED-3	nS	7.2	38.4	-31.2	-136.84
TED-4	NS	-25.5	-1.7	-23.8	175.00
TED-7-n3	n3	-3.1	20.7	-23.8	-270.45
TED-7-3N	3N	14.4	3.3	11.1	125.42
TED-7-NS	NS	0.4	10.6	-10.2	-185.45
TED-8-n3	n3	-7.6	-7.2	-0.4	5.41
TED-8-3N	3N	20.1	3.9	16.2	135.00
TED-8-NS	NS	-9.7	-24.1	14.4	-85.21
TED-5-NS	NS	-38.0	-44.2	6.2	-15.09
TED-5-3N	3N	7.7	0.9	6.8	158.14
TED-9-NS	NS	-16.5	-35.6	19.1	-73.32
TED-9-3N	3N	18.6	7.7	10.9	82.89
Average					-22.76
σ					135.77

Grouping by junction type yields table 5.11. All values of $\overline{F_{Bp-p}}$ appear dramatically higher ($\approx \times 5$ or more) than their $\overline{F_{Ap-p}}$ counterparts but at the same time diffusion-based junctions of the n-diffusion on substrate (nS) and n-diffusion on triple-well (n3) types retain the property of featuring relatively similar values. The triple-well on n-well type (3N) shows the largest $\overline{F_{Ap-p}}$ to $\overline{F_{Bp-p}}$ magnification, up by approx. 39 times.

Summary and interpretation:

Table 5.11: Average fractional difference between estimated electro-optical modulation levels (a.k.a B_{p-p}) grouped by junction type ($\overline{F_{Bp-p}}$). The last column is the number of junctions of each specific type taken into account. The fractional difference between two numbers is defined as $F_x = F(a,b) = 2\frac{a-b}{a+b}$ where 'x' sits in-lieu of (a,b).

Type	$\overline{F_{Bp-p}}$ %	**n**
nS	-143.96	2
n3	-132.52	2
3N	125.36	4
NS	-44.51	6

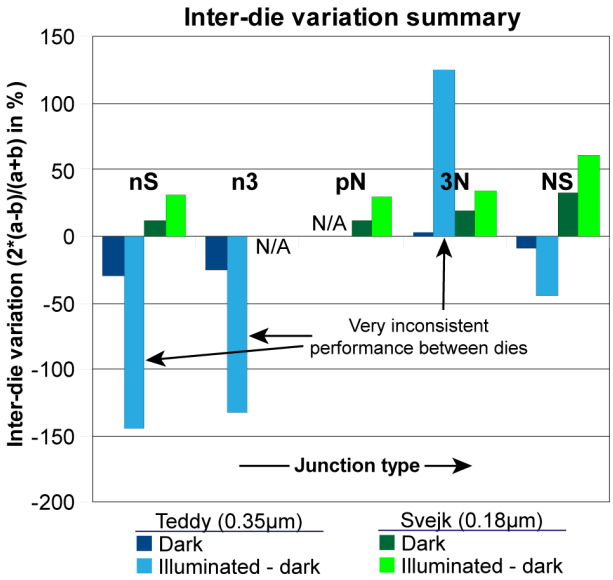

Figure 5.10: Inter-die variation in the responsivity of each junction type from each design family to our input signal. Shown are responsivities in the dark (a.k.a. 'A_{p-p}'; an estimation of pick-up noise) and in the 'illuminated minus dark' cases (a.k.a. 'B_{p-p}'; an estimation of electro-optical effects). Where no available data exists 'N/A' is displayed. Junction type legend: nS: n-diffusion on substrate. n3: n-diffusion on triple well. pN: p-diffusion on N-well. 3N: triple well on N-well. NS: N-well on substrate.

Overall, the results of our investigation into inter-die variation can be summarised in Figure 5.10. The figure shows the inter-die variation in device responses to our $8\,kHz$ input signal both in the dark (estimation of pick-up) and in the 'illuminated minus dark' case (estimation of electro-optical phenomena), sorted by junction type. Note how the variation in the estimation of the intensity of electro-optical phenomena is invariably higher (in absolute terms) compared to the estimation of pick-up noise.

Under the assumptions made for this section, we reach the conclusion that fractional differences between A_{p-p} and B_{p-p} values for both Teddy (TED) and Svejk (SVJ) device pairs are all large considering the sizes of the devices being tested. Part of these discrepancies can be assumed to arise from process variation. To that end, it would be expected that high spatial frequency effects will be averaged out by the large footprints of the devices. On the other hand, the large spread of F_{Ap-p} and F_{Bp-p} would suggest that effects comparable in scale with the size of our devices may play a considerable role. Very

low spatial frequency effects on the scale of entire dies or even entire wafers may also influence results as hinted towards by systematic offsets in modulator performance of homologous device pairs. This observation is valid both in the dark and under illumination.

The values of average fractional difference and their spreads, as obtained in the dark case differ from those obtained under illumination. This could signify that light-dependent phenomena are affected differently (and much more vigorously) by inherent variation in the constitution of each device although part of this change might be attributed to noise (the magnitudes of B_{p-p} values are generally much lower than those of A_{p-p} values and are thus much more prone to distortion through noise).

Finally we need to consider what other effects may be contributing to the observed inter-die variation. Small variations in the exact mounting of each die within its package could affect the projection of pick-up to the photodetector (akin to changing the positioning of a television antenna) thus leading to systematic offsets for each device. Without complicated analysis it is impossible to prove that these offsets will the the same or at least very similar for all devices (i.e. will be independent of device location, geometry etc.) and therefore it must be treated as a potential source of systematic offsets unique to each device. These offsets we will call 'mounting errors'. The specific contributions of noise and mounting errors are difficult to separate from those of natural inter-die variation.

Unfortunately, interesting data such as the distributions of fractional differences between pairs of homologous devices when taken across many different die pairings[14], the associated standard deviation -if the distribution is proven to be Gaussian- and any correlations between homologous device average fractional differences and location on the die (junction type-specific) cannot be usefully extracted due to the limited number of samples used.

Finally it must be noted that certain inter-die differences could potentially also be explained by 'fudge factors' such as presence of dust particles on certain devices etc. This could be offered as a possible explanation for the contrast between good Svejk design, dev. #1 (SVJ-1) and Svejk design, dev. #2 (SVJ-2) inter-die consistency and the bad Svejk design, dev. #5 (SVJ-5) and Svejk design, dev. #6 (SVJ-6) equivalent metric even though both junctions are of the same type. Nevertheless, following specific care taken to shield our dies from dust during the packaging process and given subsequent checks under a microscope we have found no obvious examples of such contamination.

5.5.2.2 Effect of device geometry on performance

In order to study the effects of geometry on modulator performance we shall utilise the following strategy: Each die will be considered individually. Svejk die #2 (SVJ2) that has been tested in both configurations I and II will be, for the purposes of analysis, considered to be a different die in these two cases although the expectation is strong that in both cases results from Svejk die #2 (SVJ2) should closely match each other qualitatively. On each die, devices will be grouped by junction type and within each junction type category, each device will yield a single value of our model function $f(A,P) = k_{areal}A + k_{fringe}P$ where $f(A,P)$ is B_{p-p} (the absolute effect of modulation so long as the assumptions described in 5.5.1 hold true), A is the extent of the areal junction component and P that of the side-wall component. k_{areal}

[14] These pairings must be done between a 'reference' device that is arbitrarily chosen as the standard and any one other of its homologous devices residing on different dies, e.g. {Svejk die #1, dev. #1 (SVJ1-1), Svejk dev. #2 dev. #1 (SVJ2-1)} vs {Svejk die #1, dev. #1 (SVJ1-1), Svejk die #3, dev. #1 (SVJ3-1)} vs {Svejk die #1, dev. #1 (SVJ1-1), Svejk die #4, dev. #1 (SVJ4-1)} etc.

and k_{fringe} are the 'junction coefficients' (we will often abbreviate this to simply 'coefficients') that we are seeking. With a minimum of two such data-points we should be able to extract both the values of k_{areal} and k_{fringe}. If more data points are available linear fits can be performed.

A caveat, however, appears with nested junctions. The best example that illustrates the problem consists of N-well on substrate (NS) type junctions. An N-well on substrate (NS) type junction can exist either as a free-standing device, or as part of a triple-well on N-well on substrate (3NS) or as part of an n-diffusion on triple well on N-well on substrate (n3NS) device. In each case the NS junction may suffer structural alterations due to the presence of other dopants in the vicinity in accordance with mechanisms akin to 'emitter push' or 'NBL' push that are often seen in CMOS manufacturing [4]. Of course, the specific manufacturing process including the order in which dopants are inserted inside the substrate will affect the extent and qualitative nature of these effects. The end result is that junctions that are of the same type may behave differently in their quality as electro-optical modulators depending on what other dopants share the same immediate vicinity with them. As a result, gathering data points from homotype junctions that originate from devices that have different sub-junction structures may prove not to be reliable.

Note: Not only electro-optical modulation, but also susceptibility to pick-up noise (measured via our A_{p-p} metric) is expected to be different between homotype test junctions that reside in devices with different overall structure (e.g. two different N-well on substrate (NS), one residing in a simple N-well on substrate (NS) type device and the other in a triple-well on N-well on substrate (3NS) type device). This will happen because the different doping properties of the compared test junctions will potentially lead to different pick-up generation efficiencies in addition to changing junction reaction to light-dependent phenomena. Moreover, the overall optical resistance that any test beam will encounter from emitter to detector will be different for the two device structures (at the very least in the area where they overlap and immediate surroundings within a few minimum corresponding design lengths).

Ninja #1 (NIN1):

Ninja 1 consists of eight devices, out of which two contained butting diffusion junctions and were proven to be unusable. Out of the remaining six devices Ninja design, dev. #2 (NIN-2) consists of nested junctions. Thus, in total, Ninja houses seven usable pn-junctions:

- N-well on substrate (NS): 4 items.
- n-diffusion on substrate (nS): 2 items.
- p-diffusion on N-well (pN): 1 item.

N-well to substrate junctions: A quick inspection reveals that out of the four existing N-well on substrate (NS) type junctions two form a degenerate pair within the context of electro-optical modulation function $f(x, y)$ because Ninja design, dev. #5 (NIN-5) and Ninja design, dev. #6 (NIN-6) are identical except for the presence of a passivation layer over Ninja design, dev. #5 (NIN-5) and its absence over Ninja design, dev. #6 (NIN-6). The number of usabe devices, however, drops further to two due to the fact that Ninja die #1, dev. #2, N-well on substrate junction (NIN1-2-NS) does not share its overall

device structure with its dev. #5 and dev. #6 counterparts[15]. Furthermore the N-well on substrate of dev. #2 (NIN-2-NS) was tested with the p-diffusion region floating.

Combining the data points from Ninja die #1, dev. #1 (NIN1-1) and Ninja die #1, dev. #5 (NIN1-5) we obtain coefficients for areal and side-wall components $-128.1\,nV/\mu m^2$ and $190.0\,nV/\mu m$ respectively. Pairing Ninja die #1, dev. #1 (NIN1-1) with Ninja die #1, dev. #2, N-well on substrate junction (NIN1-2-NS) yields considerably different values, as does pairing Ninja die #1, dev. #5 (NIN1-5) with Ninja die #1, dev. #2, N-well on substrate junction (NIN1-2-NS) (see table 5.12).

Table 5.12: Contribution of areal and side-wall junction components to electro-optical modulation (a.k.a. 'junction coefficients') extracted from various pairings of N-well on substrate (NS) -type devices on Ninja die #1 (NIN1). Dev. pair: device pair. Only identifier numbers are given for brevity. k_x,: coefficient of x.

NIN1 NS JUNCTION COEFFICIENT SUMMARY				
DIE	**DEV. PAIR**	**Type**	k_{areal}	k_{fringe}
			$nV/\mu m^2$	$nV/\mu m$
NIN1	1,5	NS	-128.1	190.0
NIN1	1,2-NS	NS	27.0	2.8
NIN1	2-NS,5	NS	131.0	-758.4

n-diffusion to substrate junctions: The pair of junctions comprising this set consists of Ninja die #1, dev. #4 (NIN1-4) and Ninja die #1, dev. #9 (NIN1-9). The extracted areal and side-wall coefficients are $126.5\,nV/\mu m^2$ and $-3.5\,nV/\mu m$.

Result summary: The resulting 'best guess' coefficients extracted from Ninja die #1 (NIN1) are summarised in table 5.13.

Table 5.13: Estimated contribution of areal and side-wall junction components to electro-optical modulation (a.k.a. 'junction coefficients') for N-well on susbstrate (NS) and n-diffusion on substrate (nS) -type junctions residing on Ninja die #1 (NIN1).

NIN1 NS JUNCTION COEFFICIENT SUMMARY				
DIE	**DEV. PAIR**	**Type**	k_{areal}	k_{fringe}
			$nV/\mu m^2$	$nV/\mu m$
NIN1	1,5	NS	-128.1	190.0
NIN1	4,9	nS	126.5	-3.5

Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) - configuration II:

The Svejk (SVJ) family offers 12 devices with a total of 15 junctions:

- N-well on substrate (NS): 8 items.
- n-diffusion on substrate (nS): 4 items.
- p-diffusion on N-well (pN): 2 items.
- Triple well on N-well (3N): 1 item.

[15]See start of this section 5.5.2.2.

'Pure' N-well to substrate junctions: Out of the eight N-well on substrate (NS) -type junctions on each Svejk (SVJ) die five originate from pure N-well on substrate (NS) devices whilst the remaining three form parts of p-diffusion on N-well on substrate (pNS) and triple-well on N-well on substrate (3NS) devices. Svejk design, dev. #7 (SVJ-7) through to Svejk design, dev. #11 (SVJ-11) are the pure N-well on substrate (NS) devices.

As can be evidenced in the design repository (appendix, section A) these devices were designed specifically to uncover finer aspects of the influence of geometry on modulator performance. Thus, we can begin our analysis by considering Svejk design, dev. #8 (SVJ-8) and Svejk design, dev. #10 (SVJ-10) that are exactly identical (they form a 'degenerate pair'), save for a difference in orientation (90 degrees difference). The Svejk die #1 (SVJ1) B_{p-p} values for these devices are $-103.9\,mV$ and $-81.7\,mV$ respectively, which corresponds to a fractional error of 23.92 %. For Svejk die #2 (SVJ2) corresponding values are $-63.6\,mV$, $-42.7\,mV$ and fractional error of 39.32 %. This means members of the degenerate pair show a $20-30\%$ variation in estimated electro-optical modulation within each die (see table 5.14).

Table 5.14: Estimated electro-optical modulation (a.k.a. B_{p-p}) levels for degenerate N-well on substrate (NS) junctions on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). These devices are called degenerate because they offer the same exposed area and perimeter to the incoming light. Their differences lie in the layout structures of their non-exposed areas. The last row shows intra-die device-to-device fractional variation according to the formula $F(a,b) = 2\frac{a-b}{a+b}$.

SVJ DEGENERATE NS DEVICE B_{p-p} VALUES			
DEVICE	**Type**	$B_{p-p}(SVJ1)$	$B_{p-p}(SVJ2)$
		mV	mV
8	NS	-103.9	-63.6
10	NS	-81.7	-42.7
F(%)		23.92	39.32

Next, we can consider the group of five pure N-well on substrate (NS) devices as a whole in order to extract junction coefficients. Mathematically we will fit a surface defined by $z = k_{areal}A + k_{fringe}P$ to the five datapoints of each die and then we will obtain pair-wise values for k_{areal} and k_{fringe} for the 'interesting' combinations of Svejk die #1 (SVJ1-) and Svejk die #2 (SVJ2-) devices #7 vs #8, #7 vs #10, #8 vs #9, #9 vs #10 and #9 vs #11. From the modulator design appendix (section A) we can show that in terms of basic cell areas SVJ devices can be ordinally ranked as: $7 < 8 = 10 < 9 < 11$ where the numbers denote device identifiers on Svejk (SVJ) dies. Thus the selected pairings correspond to device duos that feature basic cell sizes that lie adjacent to one another in this ordinal ranking. By examining the aforementioned pairings we should be able to observe whether any basic cell size-dependent non-linearities are present. For convenience we shall call pairings between Svejk design, dev. #7 (SVJ-7) and Svejk design, dev. #8 (SVJ-8) or Svejk design, dev. #10 (SVJ-10) 'ultra small - small' or 'USS' pairings. Svejk design, dev. #8 (SVJ-8) or Svejk design, dev. #10 (SVJ-10) to Svejk design, dev. #9 (SVJ-9) as 'small to medium' or 'SM', and Svejk design, dev. #9 (SVJ-9) to Svejk design, dev. #11 (SVJ-11) as 'medium to large' or 'ML'.

The surface fit for Svejk die #1, dev. #7 (SVJ1-7) through Svejk die #1, dev. #11 (SVJ1-11) data

(carried out by MATLAB) returns the areal coefficient as $-2.6\,nV/\mu m^2$ and the perimetric coefficient as $-2.8\,nV/\mu m$. The fit shows a root mean square deviation (RMSD) of $8.3468\,mV$. For Svejk die #2 (SVJ2) the corresponding coefficients are $-1.1\,nV/\mu m^2$ and $-1.7\,nV/\mu m$ and the RMSD equals $7.2205\,mV$.

Computed coefficients from 'intersting pairings' of devices residing on Svejk, die #1 (SVJ1) and Svejk die #2 (SVJ2) are summarised in table 5.15 and plotted on a Cartesian plane in Figure 5.11. Notably, corresponding coefficient values vary wildly between homologous pairings on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). We see no obvious and clear-cut monotonic trend in either areal or side-wall junction coefficient either with distance from the die corner (7 vs 8, 8 vs 9, 9 vs 10, 9 vs 11), or with basic cell size (7 vs 8, 8 vs 9, 9 vs 11 AND 7 vs 10, 9 vs 10, 9 vs 11 also see Figure 5.11). Particularly in Figure 5.11 we observe that even if we were to find the centres of gravity of each group of junction coefficients (sorted by membership to the USS, SM and ML sets) we could not be certain that any significant trend would be revealed (data spread too large). Having said that, we do note that there is a subtle hint that the centre of gravity shifts towards magnitude-smaller areal coefficients and magnitude-larger side-wall coefficients as devices get larger. More measurements of more dies are required in order to compile meaningful statistics.

Table 5.15: Contribution of areal and side-wall junction components to electro-optical modulation (a.k.a. 'junction coefficients') for various 'interesting' pairings of N-well on substrate (NS) -type devices on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Junction coefficients arising from linear fits on data from all available devices are also shown for each die for comparison. Dev. set: device pair or set based on which coefficients are extracted. Only identifier numbers are given for brevity. k_x,: coefficient of x.

SVJ1 JUNCTION COEFFICIENT SUMMARY				
DIE	**DEV. SET**	**Type**	k_{areal}	k_{fringe}
			$mV/\mu m^2$	$mV/\mu m$
SVJ1	7,8	NS	-7.2	-1.2
SVJ1	7,10	NS	-0.4	-3.3
SVJ1	8,9	NS	-2.7	-3.2
SVJ1	9,10	NS	-4.5	-1.5
SVJ1	9,11	NS	-2.0	-3.9
SVJ1	7,8,9,10,11	NS	-2.6	-2.8
SVJ2	7,8	NS	-6.8	0.4
SVJ2	7,10	NS	-1.6	-1.3
SVJ2	8,9	NS	-0.2	-2.7
SVJ2	9,10	NS	-1.9	-1.0
SVJ2	9,11	NS	-1.0	-1.8
SVJ2	7,8,9,10,11	NS	-1.1	-1.7

n-diffusion to substrate junctions: Out of the four n-diffusion on substrate (nS) junctions residing on each Svejk (SVJ) die two of them, Svejk design, dev. #1 (SVJ-1) and Svejk design, dev. #2 (SVJ-2), form a degenerate pair (same area, same perimeter) much like Svejk design, dev. #8 (SVJ-8) and Svejk design, dev. #10 (SVJ-10) in the N-well on substrate (NS) family. Whereas Svejk design, dev. #8 (SVJ-8) and Svejk design, dev. #10 (SVJ-10) only differed in orientation, the Svejk design, dev. #1 (SVJ-1), Svejk design, dev. #2 (SVJ-2) pair plays upon the different extents of their respective diffusion regions underneath the metal mask. In other words, they both present the same areal element and same (zero-length) perimetric element to the incoming illumination, but in Svejk design, dev. #2 (SVJ-2) the

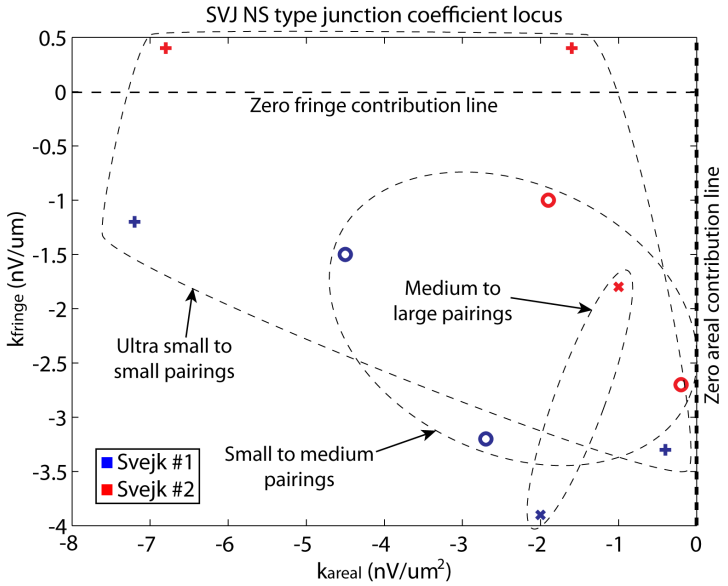

Figure 5.11: Estimated areal and side-wall junction component contribution to electro-optical modulation (a.k.a. 'junction coefficients') locus for various 'interesting' pairings between pure N-well on susbtrate (NS) devices on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Each colour represents a different die. Blue: Svejk die #1 (SVJ1). Red: svejk die #2 (SVJ2). Each symbol represents pairings between devices featuring different basic cell sizes. +: ultra-small to small. o: small to medium. x: medium to large. For more details see text.

junction perimeter lies farther away from the illuminated 'core' of the device than in Svejk design, dev. #1 (SVJ-1) -see Figure 5.12-. The other devices all form non-degenerate pairings amongst them.

We begin by examining the degenerate Svejk design, dev. #1 (SVJ-1), Svejk design, dev. #2 (SVJ-2) pairing. The estimated electro-optical modulation intensity (B_{p-p}) values we obtain for Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) are given in table 5.16. We notice how the members of our degenerate pair seem to behave in fairly similar ways ($< 6\%$ difference between members of the pair on each die). This indicates that the extent of the diffusion region underneath the metal mask makes little difference to modulator performance, at least when the closest diffusion region edge is a few microns away from the exposed area (note that the n-diffusion on substrate type junction in this technology has a minimum design length of approx. 1/5 of a micron).

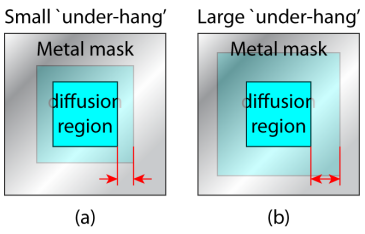

Figure 5.12: Structural differences between (a) Svejk design, dev. # 1 (SVJ-1) and (b) Svejk design, dev. # 2 (SVJ-2). The metal mask will, of course, be opaque to electro-magnetic radiation, but is shown as semi-transparent here in order to offer a sense of depth. The amount by which the diffusion region of each device extends below the metal mask (the 'under-hang') is the main difference as evidenced by the red arrows.

Table 5.16: Estimated electro-optical modulation (a.k.a. B_{p-p}) levels for degenerate n-diffusion on substrate (nS) junctions on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). These devices are called degenerate because they offer the same exposed area and perimeter to the incoming light. Their differences lie in the layout structures of their non-exposed areas. The last row shows intra-die device-to-device fractional variation according to the formula $F(a,b) = 2\frac{a-b}{a+b}$.

SVJ DEGENERATE nS DEVICE B_{p-p} VALUES			
DEVICE	**Type**	$B_{p-p}(SVJ1)$	$B_{p-p}(SVJ2)$
		mV	mV
1	nS	-99.1	-87.6
2	nS	-105.1	-91.0
F(%)		-5.88	-3.81

Next, we examine the surface fit to all four n-diffusion on substrate (nS) type devices. MATLAB fitting yields coefficient values of $-6.4\,nV/\mu m^2$ and $6.2\,nV/\mu m$ for Svejk die #1 (SVJ1) and $-4.6\,nV/\mu m^2$ and $4.8\,nV/\mu m$ for Svejk die #2 (SVJ2). The RMSD values were equal to $24.81\,mV$ for Svejk die #1 (SVJ1) and $31.26\,mV$ for Svejk die #2 (SVJ2). Note: since three out of the four devices have zero-length exposed perimeters, the quality of the fit might have been severely compromised at least in what concerns the side-wall junction coefficient. As such the values given here should be considered with caution.

p-diffusion to N-well junctions: Only two p-diffusion on N-well (pN) type junctions were designed on the Svejk (SVJ) family. These form a single pair and therefore yield a pair of junction coefficients for each die. The resulting coefficients are $-42.7\,nV/\mu m^2$ and $60.8\,nV/\mu m$ for Svejk die #1 (SVJ1) and $-31.9\,nV/\mu m^2$ and $45.7\,nV/\mu m$ for Svejk die #2 (SVJ2). Notably, one of the devices forming the pair of Svejk design, devices #4 (SVJ-4) has none of its perimeter exposed to the incoming illumination.

N-well on substrate devices under p-diffusion (p-diffusion floating): The same devices that offer the p-diffusion on N-well (pN) junctions (Svejk design, dev. #3 (SVJ-3) and Svejk design, dev. #4 (SVJ-4)) also allow us to investigate how N-wells behave when covered by p-type diffusion. This was not the primary objective of these designs, however. Unfortunately, both N-well on substrate (NS) components of Svejk design, dev. #3 (SVJ-3) and Svejk design, dev. #4 (SVJ-4) only show an areal component to the illumination. This only allows the study of the areal component and presents at the same time a conundrum. Table 5.17 summarises the estimated electro-optical modulation levels (B_{p-p} values) for Svejk design, dev. #3 (SVJ-3) and Svejk design, dev. #4 (SVJ-4), N-well on substrate (NS) components for both Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Important note: the front-ends of Svejk design, dev. #3 (SVJ-3) and Svejk design, dev. #4 (SVJ-4) are identical. They only differ in their metal masks.

Table 5.17 makes it very clear that the dramatic difference in exposed areas is not reflected by a similarly dramatic change in B_{p-p} value. On both Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) B_{p-p} values for both devices #3 and #4 are similar, which means that contrary to expectation, no device appears to be an unequivocally stronger modulator.

The p-diffusion on N-well on substrate (pNS) combination: For Svejk design, dev. #3 (SVJ-3) and Svejk design, dev. #4 (SVJ-4) and the Svejk design (SVJ), the host PCB offers the possibility of examining the entire p-diffusion on N-well on substrate (pNS) structure as a single pn-junction unit where the p-side consists of the substrate and the p-diffusion region shorted together and the n-side consists of

Table 5.17: Estimated electro-optical modulation levels (B_{p-p} values) for Svejk die #1, dev. #3 (SVJ1-3), Svejk die #1, dev. #4 (SVJ1-4), Svejk die #2, dev. #3 (SVJ2-3) and Svejk die #2, dev. #4 (SVJ2-4) devices, N-well on substrate (NS) type components. The exposed area of each NS junction is also shown.

SVJ pN PAIRING B_{p-p} VALUES				
DEVICE	**Type**	$B_{p-p}(SVJ1)$	$B_{p-p}(SVJ2)$	Area
		mV	mV	μm^2
3	NS	-139.6	-62.3	31819
4	NS	-142.2	-61.4	6757

the N-well region. Ideally the p-diffusion on N-well on substrate (pNS) modulation performance will be the linear sum of the p-diffusion on N-well (pN) and N-well on substrate (NS) components. Table 5.18 shows clearly that this does not occur.

Table 5.18: Estimated electro-optical modulation levels (B_{p-p} values) for Svejk design, dev. #3 (SVJ-3), Svejk design, dev. #4 (SVJ-4) devices on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Values are given for both individual pn-junctions within the device and in the case where the substrate and p-diffusion regions are shorted together ('pNS' case).

SVJ pN PAIRING B_{p-p} VALUES			
DEVICE	**Configuration**	$B_{p-p}(SVJ1)$	$B_{p-p}(SVJ2)$
		mV	mV
3	pN	-281.3	-207.3
3	NS	-139.6	-62.3
3	pNS	-94.4	-30.5
4	pN	-288.2	-215.6
4	NS	-142.2	-61.4
4	pNS	-92.8	-31.9

Result summary: The 'best guess' junction coefficients extracted from Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) in configuration II are summarised in table 5.19.

Svejk die #2 (SVJ2) - configuration I:

In configuration I the results of Svejk die #2 (SVJ2) were expected from the beginning to be quantitatively different compared to those arising from test in configuration II. However, qualitative results are still expected to follow similar trends on both configurations. Since most of the analysis is directly related to the one carried out for configuration II only a 'best guess' junction coefficient table will be provided along with some comments on the resulting data.

Results: Table 5.20 summarises the fitted/computed areal and side-wall junction coefficients. The errors associated with the fits are equal to $93.5\,pV$ for the N-well on substrate (NS) type junction fit and $218.1\,pV$ for the n-diffusion on substrate (nS) equivalent. Of note is the fact that in terms of absolute values N-well on substrate (NS) type junctions are the weakest with p-diffusion on N-well (pN) being the best and n-diffusion on substrate (nS) performing between these two extremes. This applies to both areal and side-wall coefficients. Areal coefficients from all device types all agree in sign, which isn't the case for their side-wall counterparts. Finally, the absolute value of the k_{areal} for N-well on substrate

Table 5.19: 'Best guess' areal and side-wall junction electro-optical modulation contribution coefficients (a.k.a. 'junction coefficients') for all types of junctions residing on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). These are based on results from test configuration II. DEV. SET: The set of devices used to perform the linear fit that yielded the junction coefficients.

JUNCTION COEFFICIENT SUMMARY				
DIE	**DEV. SET**	**Type**	k_{areal}	k_{fringe}
			$nV/\mu m^2$	$nV/\mu m$
SVJ1	7,8,9,10,11	NS	-2.6	-2.8
SVJ2	7,8,9,10,11	NS	-1.1	-1.7
SVJ1	1,2,5,6	nS	-6.4	6.2
SVJ2	1,2,5,6	nS	-4.6	4.8
SVJ1	3,4	pN	-42.7	60.8
SVJ2	3,4	pN	-31.9	45.7

(NS) junctions seems to be much smaller than all other junction coefficients. Note changed units of k_X which are now in pV as opposed to nV per unit area or perimeter accordingly.

Table 5.20: 'Best guess' areal and side-wall junction electro-optical modulation contribution coefficients (a.k.a. 'junction coefficients') for all types of junctions residing on Svejk die #2 (SVJ2). These are based on results from test configuration I. DEV. SET: The set of devices used to perform the linear fit that yielded the junction coefficients.

JUNCTION COEFFICIENT SUMMARY				
DIE	**DEV. SET**	**Type**	k_{areal}	k_{fringe}
			$pV/\mu m^2$	$pV/\mu m$
SVJ2	7,8,9,10,11	NS	-0.9	-21.1
SVJ2	1,2,5,6	nS	-31.9	40.9
SVJ2	3,4	pN	-60.7	152.3

Result summary: Configuration I is a qualitatively similar, but quantitatively massively altered set-up compared to configuration II. Thus, we can expect that any observations made on the topics of consistency of results between dies, levels of uncertainty etc. that applied to configuration II will not necessarily also hold in configuration I. This is problematic as there is no data that can cross-validate information extracted from Svejk die #2 (SVJ2) in configuration I. Instead, data obtained from configuration I measurements must be related to equivalent data from configuration II in search of -ideally- a single scaling factor that will match the results of each configuration to one another.

The obvious correlation should be between measurements on Svejk die #2 (SVJ2) in both configurations. Table 5.21 summarises the ratios of k_{areal} and k_{fringe} in the form of Svejk die #2 (SVJ2) configuration II over Svejk die #2 (SVJ2) configuration I. N-well on substrate (NS) junctions show a rather sizable discrepancy in their scaling factors (areal over side-wall scaling factor ratio of approx. 14.7). n-diffusion on substrate (nS) type junctions behave more or less as expected with a fairly consistent scaling factor (although still with a scaling factor fractional difference of approx. 20.5%[16]). Finally, p-diffusion on N-well (pN) type junctions behave similarly to their n-diffusion on substrate (nS) counterparts, but with an increased scaling factor fractional difference of approx. 54.6%.

[16] Where fractional difference, as usual, is defined by $F(a,b) = 2\frac{a-b}{a+b}$.

Table 5.21: Ratios of 'best guess' areal and side-wall junction electro-optical modulation contribution coefficients of Svejk die #2 (SVJ2) configuration II over Svejk die #2 (SVJ2) configuration I. Also shown are the ratios of scaling factor ratios and the fractional difference between them ($F(k_{areal}/k_{fringe})$).

SVJ2 COEFFICIENT RATIO BETWEEN TEST CONFIGS						
DIE	**DEV. PAIR**	**Type**	k_{areal} **ratio**	k_{fringe} **ratio**	$\frac{k_{areal}ratio}{k_{fringe}ratio}$	$F(k_{areal}/k_{fringe})$
SVJ2	7,8,9,10,11	NS	1182.8	80.53	14.69	1.75
SVJ2	1,2,5,6	nS	144.2	117.36	1.23	0.21
SVJ2	3,4	pN	525.54	300.07	1.75	0.55

The fractional difference data seems to indicate that either something fundamental changes when switching from one configuration to the other or results from one of the configurations or both are corrupted by noise. Qualitatively, all behaviours are loosely preserved but the absolute B_{p-p} values extracted from configuration I are significantly smaller than their configuration II counterparts. This immediately renders all configuration I results much more vulnerable to noise at least from all the electronics downstream of the photodetector; electronics whose noise response shouldn't be affected in any particularly direct way by the level of illumination or sensed A_{p-p} value (a measure of pick-up noise). Furthermore, the coefficients of N-well on substrate (NS) type junctions (the ones showing the worst scaling factor discrepancies) are the ones which are smallest in magnitude and therefore most vulnerable to noise. Robust evidence of direct correspondence between our two test configurations can only be found if more measurements are taken with more dies, yet the above-stated observations hint towards the possibility that the inconsistencies between results taken in configurations I and II on Svejk die #2 (SVJ2) can be at least largely attributed to electrical, and possibly also optical noise.

Teddy die #5 (TED5) and Teddy die #6 (TED6):

The Teddy (TED) design family dies feature eight devices consisting of a total of 14 pn-junctions:

- N-well on substrate (NS): 6 items.
- Triple well on N-well (3N): 4 items.
- n-diffusion on triple well (n3): 2 items.
- n-diffusion on substrate (nS): 2 items.

N-well to substrate junctions: Out of the six N-well on substrate (NS) type junctions available, we have to discriminate between a pair of pure N-well on substrate (NS) junctions (Teddy design, dev. #1 (TED-1) and Teddy design, dev. #4 (TED-4)), a pair where the N-well hosts a triple well (Teddy design, dev. #5 (TED-5) and Teddy design, dev. #9 (TED-9)) and a pair where the N-well hosts a triple-well, which itself hosts an n-diffusion region (Teddy design, dev. #7 (TED-7) and Teddy design, dev. #8 (TED-8)). We shall consider only the three device pairings corresponding to N-well on substrate (NS) junctions that are components of the same overall device type. In all cases any 'upstream' doped regions (triple wells, n-diffusion) have been left floating.

Table 5.22 summarises the extracted junction coefficients for all N-well on substrate (NS) junction pairings. Notably, there seems to be very little consistency between Teddy die #5 (TED5) and Teddy die #6 (TED6) in terms of computed junction coefficients although five out of six coefficients of each die agree in sign. In the case where corresponding coefficients do not agree in sign (pairing {5,9}, areal coefficient) the value contributed by Teddy die #5 (TED5) is very close to zero and may have been affected by noise.

Finally, for the n-diffusion n triple well, on N-well on substrate (n3NS) case the values obtained for Teddy die #6 (TED6) seem to be a scaled up version of the values obtained for Teddy die #5 (TED5) ($\times$3.37 for the areal coefficient and $\times$3.46 for the side-wall coefficient).

Table 5.22: Summary of junction coefficients as extracted from pairings between N-well on substrate (NS) junctions residing on homotype devices.

JUNCTION COEFFICIENT SUMMARY					
DIE	**DEV. PAIR**	**Type**	**Host type***	k_{areal}	k_{fringe}
				$nV/\mu m^2$	$nV/\mu m$
TED5	1,4	NS	NS	-1.4	0.5
TED5	5,9	NS	3NS	-0.1	-0.8
TED5	7,8	NS	n3NS	-3.0	26.5
TED6	1,4	NS	NS	-0.2	0.3
TED6	5,9	NS	3NS	1.3	-17.3
TED6	7,8	NS	n3NS	-10.1	91.7

*The column marked 'host type' denotes the full specification of the devices being tested.

Triple-well to N-well junctions: Out of the four triple-well on N-well (3N) junctions present on Teddy (TED) dies, two are part of triple well on N-well on substrate (3NS) devices whilst the other two are components of n-diffusion on triple well on N-well on substrate (n3NS) devices. The junction coefficients will be extracted from pairs of triple well on N-well (3N) junctions that reside on homostructure devices only. If n-diffusion regions were present they were left floating for these experiments.

Table 5.23 displays the extracted junction coefficients. All corresponding coefficients between Teddy die #5 (TED5) and Teddy die #6 (TED6) agree in sign. Junctions residing in triple well on N-well on substrate (3NS) devices show higher coefficients in Teddy die #5 (TED5) and junctions residing in n-diffusion on triple well on N-well on substrate (n3NS) devices showing higher coefficients in Teddy die #6 (TED6).

n-diffusion on triple-well junctions: The lone pair of n-diffusion on triple well (n3) type junctions present on Teddy (TED) dies are members of the nested-junction n-diffusion on triple well on N-well on substrate (n3NS) Teddy design, dev. #7 (TED-7) and Teddy design, dev. #8 (TED-8). Theoretically n-diffusion on triple well (n3) junctions can also be manufactured in pure n-diffusion on triple-well (n3) devices where the triple-well is simply allowed to contact the p-type substrate in an ohmic fashion, but such devices have not been manufactured for this project.

The extracted coefficients are $0.01\,nV/\mu m^2$ (areal) and $-0.49\,nV/\mu m$ (fringe) for Teddy die #5 (TED5) and $0.6\,nV/\mu m^2$ and $-1.1\,nV/\mu m$ for Teddy die #6 (TED6) respectively. Corresponding coefficients between Teddy die #5 (TED5) and Teddy die #6 (TED6) agree in sign but the differences

Table 5.23: Summary of junction coefficients as extracted from pairings between triple well on N-well (3N) junctions residing on homotype devices.

JUNCTION COEFFICIENT SUMMARY					
DIE	**DEV. PAIR**	**Type**	**Host type***	k_{areal}	k_{fringe}
				$nV/\mu m^2$	$nV/\mu m$
TED5	5,9	3N	3NS	0.44	-0.09
TED5	7,8	3N	n3NS	0.17	0.50
TED6	5,9	3N	3NS	0.19	-0.06
TED6	7,8	3N	n3NS	0.43	0.77

*The column marked 'host type' denotes the full specification of the devices being tested.

between them are rather large (factor of approx. 60 for areal coefficient and approx. 2.25 for fringe).

n-diffusion on substrate junctions: The lone pair of n-diffusion on substrate (nS) devices on Teddy (TED) designs consists of Teddy design, dev. #2 (TED-2) and Teddy design, dev. #3 (TED-3). The extracted coefficients are $0.0359\,nV/\mu m^2$ (areal) and $0.0012\,nV/\mu m$ (fringe) for Teddy die #5 (TED5) and $0.1893\,nV/\mu m^2$ and $0.0281\,nV/\mu m$ for Teddy die #6 (TED6) respectively. These are very small values and maintain high inconsistencies between dies. The coefficients on Teddy die #6 (TED6) are far higher than on Teddy die #5 (TED5).

Conclusions and summary:

The core information describing the effects of geometry on modulator performance consists of the so-called 'junction coefficients'; the areal coefficient k_{areal} and the side-wall coefficient k_{fringe}. Values for these coefficients have been extracted from each device by use of different methods depending on the number of devices involved in the extraction process:

- Single device: This is a special case applying only to Ninja die #1, dev. #6 (NIN1-6). Since it is identical to Ninja die #1, dev. #5 (NIN1-5) in all but the absence of passivation coating over it its coefficients were calculated by extrapolation from Ninja die #1, dev. #5 (NIN1-5).
- Paired devices: When only two devices were available the linear system of the form $z = Kx$ was solved where z represents the B_{p-p} values corresponding to each device, K is a square 2×2 matrix that holds geometrical data about them and x is the vector of areal and side-wall junction coefficients to be found. MATLAB was used to obtain the actual results.
- Devices grouped in numbers greater than two: When such case arises, the linear system to be solved $z = Kx$ becomes over-defined. We can still extract junction coefficients, but the extracted values will be fits rather than specific values. MATLAB physically carried out these fitting operations. These fits also come with a root-mean-square deviation values attached to them.

The result of these operations was a table consisting of all junction coefficients that our array of devices grants access to: table 5.24. Table 5.25 summarises fit error data and sets it against the backdrop of the A_{p-p} and B_{p-p} values featured by the fitted devices.

Table 5.24: General summary of all extracted normalised, areal and side-wall junction, electro-optical modulation contribution coefficients (a.k.a. 'junction coefficients') from all dies and device groupings.

JUNCTION COEFFICIENT SUMMARY					
DIE	**DEV. PAIR**	**Type**	**Host type**	k_{areal}	k_{fringe}
				$nV/\mu m^2$	$nV/\mu m$
NIN1	1,5	NS	NS	-128.1	190.0
	6	NS	NS	-55.2	81.9
	4,9	nS	nS	126.5	-3.5
SVJ1	7,8,9,10,11	NS	NS	-2.6	-2.8
	1,2,5,6	nS	nS	-6.4	6.2
	3,4	pN	pN	-42.7	60.8
SVJ2 (II)	7,8,9,10,11	NS	NS	-1.1	-1.7
	1,2,5,6	nS	nS	-4.6	4.8
	3,4	pN	pN	-31.9	45.7
SVJ2 (I)	7,8,9,10,11	NS	NS	-0.9	-21.1
	1,2,5,6	nS	nS	-31.9	40.9
	3,4	pN	pN	-60.7	152.3
TED5	1,4	NS	NS	-1.4	0.5
	5,9	NS	3NS	-0.1	-0.8
	7,8	NS	n3NS	-3.0	26.5
	5,9	3N	3NS	0.44	-0.09
	7,8	3N	n3NS	0.17	0.50
	7,8	n3	n3NS	0.01	-0.49
	2,3	nS	nS	0.04	0.00
TED6	1,4	NS	NS	-0.2	0.3
	5,9	NS	3NS	1.3	-17.3
	7,8	NS	n3NS	-10.1	91.7
	5,9	3N	3NS	0.19	-0.06
	7,8	3N	n3NS	0.43	0.77
	7,8	n3	n3NS	0.60	-1.10
	2,3	nS	nS	0.19	0.03

DEV. PAIR: grouped device identifier numbers. Host type: full structure of devices to which the grouped junctions belong. k_{areal}: areal junction coefficient describing B_{p-p} magnitude per unit area. k_{fringe}: similar to k_{areal}, but relating B_{p-p} magnitude to unit side-wall length. For Svejk die #2 (SVJ2) the configuration type is also given in parentheses in the first column. Note: the units for Svejk die #2 (SVJ2) in configuration I are in $pV/\mu m^2$ and $pV/\mu m$ respectively, but that is not explicitly shown by the table headers.

Table 5.25: Summary of fit errors for all multi-device groupings considered when extracting electro-optical modulation capability junction coefficients. A_{p-p} and B_{p-p} value ranges for the test devices are also given. B_{p-p} value range serves as a comparison to the magnitude of the RMSD. A_{p-p} value ranges on the other hand indicate background activity and serve as a comparison to B_{p-p} values.

FIT ERROR SUMMARY TABLE				
DIE	**DEV. GROUP**	A_{p-p} **range (dark)**	B_{p-p} **range**	**Fit error**
		mV	mV	mV
SVJ1	7,8,9,10,11	[245.1, 1024.3]	[-103.9, -76.9]	8.35
SVJ1	1,2,5,6	[263.5, 1884.3]	[-110.6, -93.3]	24.81
SVJ2 (II)	7,8,9,10,11	[183.5, 856.1]	[-63.6, -39.6]	7.22
SVJ2 (II)	1,2,5,6	[221.7, 1855.1]	[-91.0, -59.9]	31.26
SVJ2 (I)	7,8,9,10,11	[13.18, 15.56]	[-0.58, -0.15]	0.09
SVJ2 ()	1,2,5,6	[12.84, 18.54]	[-0.68, -0.22]	0.22

From these results we have already drawn a number of conclusions. Some of them arise from generic observations made by inspecting the entire data set while others arise from observations of the behaviour of certain dies or devices:

Junction coefficient signs: Interestingly and quite contrary to expectation we have observed junction coefficients in every technology featuring both positive and negative signs. The expectation was that all coefficient signs would be the same because electro-optical effects take place around the entire metallurgical surface of each junction and has similar effects throughout it (higher reverse bias voltage implies fewer light amplitude losses and vice versa). There is no solid explanation as to why that should occur, but the signs of coefficients show good consistency between different dies thus suggesting that the phenomenon giving rise to coefficient signs is not a random uncertainty factor.

A large number of cases (specific device groupings) have shown that their areal and side-wall coefficients are opposite in sign (e.g. Ninja die #1, devices #1 and #5 (NIN1-1,5) or Svejk die #1, devices #3 and #4 (SVJ1-3,4)). This may simply be the result of the fact that the presence of a side-wall degrades areal junction performance in its vicinity to the degree that the side-wall's own contribution is outweighed by this 'damage'. In that the side-walls make a nominally negative contribution to electro-optical modulation even though their presence still adds to the overall modulation exercised by the device. This, explanation seems to conform to theoretical expectations (namely: that every piece of pn-junction contributes towards modulation regardless of side, type or orientation).

Junction coefficient magnitudes: The magnitudes of all junction coefficients seem to vary wildly with technology (see 5.5.2.4) and junction type (see 5.5.2.3 later), ranging from tens of pV per unit area or unit length to over one hundred mV per unit area or length. Inter-die consistency is maintained relatively well at sign level, but begins to show large variations when we attempt to quantify the inter-die discrepancies. This seems to be particularly the case wherever the extracted coefficients are very low in magnitude. This could be attributed to noise components that do not scale with the factors that give rise to junction coefficients, notably purely electrical noise added throughout the set-up. Nevertheless, in some cases, such as with dies Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) (configuration II) it seems possible to discern clear differences in junction coefficients between junction types because the inter-die consistency allows that to happen (see 5.5.2.3).

Areal and fringe coefficients extracted from groups or pairs of devices do not seem to show any solid trend as to which coefficient value is higher. On some occasions the areal coefficient is larger in magnitude (e.g. Ninja die #1 devices #4 and #9 (NIN1-4,9)), other times it is smaller (e.g. Svejk die #2, devices #3 and #4 (SVJ2-3,4) -configuration I and II-) and other times similar in magnitude (e.g. Svejk die #1, devices #7, #8, #9, #10 and #11 (SVJ1-7,8,9,10,11)). This indicates that if there are solid trends in the relationship between areal and side-wall junction coefficients they may be significantly different depending on the type of junction being used and/or on the manufacturing technology. This is expected given the fundamentally different relations between side-wall and areal components for different junction types. We will examine this issue in more detail in 5.5.2.3.

Note: admittedly comparing absolute magnitudes of coefficients that measure B_{p-p} per unit area with coefficients that measure B_{p-p} per unit perimeter length is a rather arbitrary method of comparison. However, in this specific case the numbers allow us to illustrate the point. A much more appropriate method of examining the relation between areal and fringe junction coefficients might have been to use areal/side-wall coefficient magnitude ratios and see how the population of ratios varies within our device sample space (the set of all individual devices manufactured for this project).

Fitting strategy: Wherever fitted data is presented, it corresponds to a fitting against a linear model which corresponds to a function of the form $F(A, P) = k_{areal}A + k_{fringe}P$, with A and P standing for area and perimeter respectively and $F(A, P)$ representing B_{p-p}. This model may not necessarily be an accurate description of everything that B_{p-p} data represents. Having so little data at our disposal renders the task of validating the fitting strategy very hard. The behaviour of N-well on substrate (NS) in p-diffusion on N-well on substrate (pNS) devices on Svejk (SVJ) dies strengthens the suspicion that the simple linear model proposed may not be a good description of the observed data set.

Blurring of difference between side-wall and areal junction components: In devices with small basic cells, where designed features are of the same order as the minimum allowable features, it may be possible that a differentiation between side-wall and areal components is not meaningful. In turn, possible blur between areal and side-wall junction components may have led to significant degradation of junction coefficient extracted values. In the absence of knowledge about specific doping profiles that claim can not be verified but rather speculated upon. In any CMOS device, the actual doping concentration as a function of location depends on the entire manufacturing process (particularly the order, duration and temperature of high-temperature manufacturing steps) that can at best be approximated by crude analytical models or meticulous numerical simulation.

This affects a very large number of devices as evidenced by the small basic cell sizes we have chosen for them (see design repository in the appendix A for exact numbers). As an example we can give the case of the pairings between Svejk design, devices #7 through #11 (SVJ-7,8,0,10,11) seen earlier in this section, where the possibility of this blurring effect may explain why the centre of gravity of of each 'grouping' of pairings (see figure 5.11) tends to shift towards magnitude-smaller areal coefficients and magnitude-larger side-wall coefficients as the devices involved get larger: with larger devices we can much more easily split our junctions into clear-cut areal and side-wall regions and therefore the computed coefficients are closer to their true values.

Inter-junction interaction and homotype junctions in heterotype devices: It is a known fact that when dopants of different types are forced to occupy similar volumes of Silicon they interact with each other and cause doping profiles to develop differently than if each dopant was introduced into the Silicon

individually[17]. In areal junction components this is a non-engineerable interaction but in side-walls the distance between the edge of one doped region (e.g. N-well) from a nearby counter-doped region (e.g. a triple-well) can be controlled. Thus side-wall junction component doping characteristics will vary in accordance to how other dopants are placed in their immediate area and therefore modulator performance will change by varying the spatial pattern of doped regions via layout design. We suspect this phenomenon to have affected devices where many pn-junction side-walls lie in close proximity to one another with particular severity, such as is the case for Teddy design, devices #7 and #8 (TED-7 and TED-8).

Note: In terms of their areal components, test junctions will be likely affected by the presence of other doped species in the vicinity as was mentioned before. However, due to lack of control over doping profile in the direction perpendicular to the surface of the die these effects should be fairly consistent, particularly in areas that lie far away from any doped region edges (i.e. in the absence of corner and edge effects). Such areas can be found, for example, towards the middle of extensive, square devices.

As a result of such mechanisms, junctions of the same type but belonging to devices with different structures seem to behave significantly differently. The presence of other dopant atoms in the immediate area does seem to affect results. This is indicated by the fact that inter-die variation between junctions of the same type belonging to devices of the same type seems to be lower than the variation between devices of the same type but belonging to different device structures on the same die. Good examples of affected devices are yielded by the various 'flavours' of N-well on substrate devices residing on the UMC13 dies, the Teddies. This is considered in more detail in 5.5.2.3.

Intra-die variation: There seems to be significant intra-die variation throughout our test devices[18], i.e. similar devices on different locations of the same die behave rather differently. Given the large size of these devices we have arrived at the conclusion that under the strict assumption that B_{p-p} values accurately reflect electro-optical modulation, the observed intra-die discrepancies can be due to a handful of factors: a) Low spatial frequency manufacturing variations. b) Chip warping. c) Asymmetry in side-wall junctions that run in different directions. d) Non-uniform photodetector sensitivity, i.e. light falling on different areas of the photodetector generates different photocurrents. e) Disturbances in the light source between measurements. There was an attempt to mitigate the last of these factors by taking all measurements in quick succession.

A lot of these possible sources of intra-die variation could be tested for on Svejk (SVJ) N-well on substrate (NS) type devices. Closer examination of Svejk design,. dev. #7 (SVJ-7) through Svejk design,. dev. #11 (SVJ-11) data hinted towards no one of the above as an obvious dominant factor. Examination of the Svejk design,. dev. #7 (SVJ-7) vs Svejk design,. dev. #8 (SVJ-8) pairing in relation to all others within the Svejk design,. dev. #7 (SVJ-7) through Svejk design,. dev. #11 (SVJ-11) group revealed unusually high areal coefficient values. Since Svejk design,. dev. #7 (SVJ-7) sits at the very corner of the Svejk (SVJ) die that outlier may be potentially attributed preferentially to chip warping.

Effects of passivation layer: The presence or absence of a passivation layer does seem to affect B_{p-p} values as evidenced by the performances of Ninja design, devices #5 and #6 (NIN-5 and NIN-6). An alternative explanation of the observed discrepancies in the performances of these devices would involve

[17] As would be seen when comparing a pure N-well on substrate (NS), a pure p-diffusion on N-well (pN) and a p-diffusion on N-well on substrate (pNS) structure residing on the same die and therefore all of them having been subjected to the exact same manufacturing process, including all the high temperature steps.

[18] For which we can define inter-die variations, i.e. those residing on dies, multiple instances of which have been tested.

intra-die variation factors causing some or all of the observed differences between devices technically discriminable only in the presence/absence of their passivation coating.

Floating node effects: We have considered the possibility that leaving nodes floating during measurements may distort results obtained from the junction under test. The precise connectivity between different semiconductor regions can affect measured photocurrents and depletion region widths as a function of bias voltage.

5.5.2.3 Effect of device type on performance

Having built tables of junction coefficients extracted from a variety of junction groupings spanning many junction types, dies and manufacturing technologies we can begin to examine how these coefficiets vary between junction types. A way of visualising the results will be to plot junction coefficients on a Cartesian plane with the x-axis representing the areal coefficient and the y-axis representing the side-wall coefficient. We shall call such plots 'coefficograms'. Throughout this section we shall show coefficograms for each technology. If more than one die is involved in the construction of the coefficogram data acquired from the devices groupings on each die will be presented as separate data points (rather than pooling them by e.g. averaging).

Ninja die #1 (NIN1):

We begin areal and side-wall coefficient analysis by examining Ninja die #1 (NIN1). Three sets of junction coefficients are available from Ninja die #1 (NIN1), but one of them (Ninja die #1, dev. #6 (NIN1-6)) is based on extrapolation from another (Ninja die #1, devices #1 and #5 (NIN1-1,5)) and concerns a device that has been manufactured without a passivation cover. Thus, we shall ignore this data point and compare the remaining two. The junction coefficient scatter chart for Ninja die #1 (NIN1) can be seen in Figure 5.13.

We notice that the separation between the N-well on substrate (NS) and n-diffusion on substrate (nS) coefficients is large compared to their distance from the origin (0,0), which represents the totally unresponsive element (no modulation contribution from either junction component). We can also see that N-well on substrate (NS) junctions are about as responsive as n-diffusion on substrate (nS) junctions in terms of areal coefficients, but much more responsive in terms of side-wall coefficients. Unfortunately, the side-wall and areal components of N-well on substrate (NS) junctions seem to counteract each other's influence on overall B_{p-p} output, which probably means that the presence of side-walls hugely diminishes the effectiveness of neighbouring areal junction territory in terms of modulation. Thus, n-diffusion on substrate (nS) junctions seem more appropriate for use in modulator designs for technology AMS35 for high perimeter/area ratios whilst N-wells may still be competitive for low perimeter/area ratios.

Naturally, these numbers can not be taken at face value given that a) they arise from simple device pairings rather than fits and b) there are no other available data-points that can at least provide us with some estimate of spread. Fortunately, however, dies in technologies IBM18 and UMC13 (Svejks and Teddies respectively) will provide two data-points for each junction type and some of those will be fits coming from multiple devices. Should junction coefficients in both those design families show reasonably small spreads we can suspect that in AMS35 too the spread of junction coefficients is likely to be comparably small. This, of course, is by no means proof and further study is required to ascertain the exact behaviour of AMS35 dies.

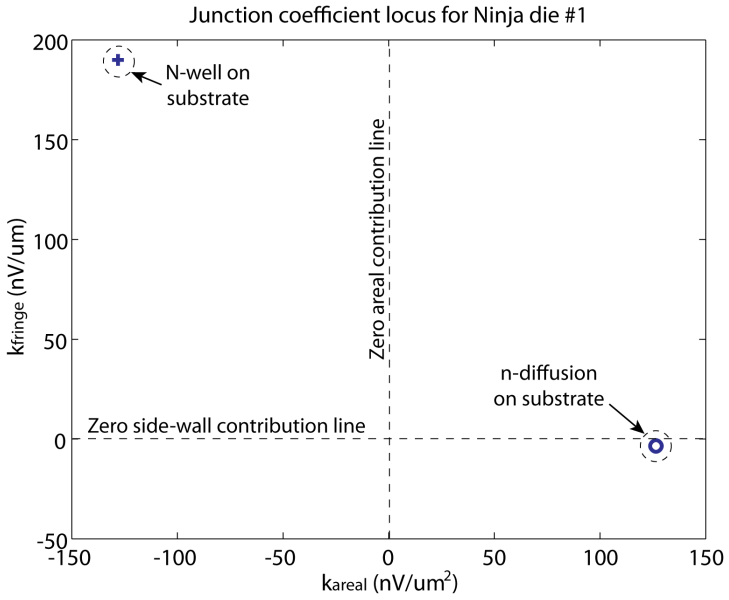

Figure 5.13: Normalised electro-optical modulation contributions of areal and side-wall junction components, pooled by junction type (the 'junction coefficients'), for Ninja die #1 (NIN1). Each symbol represents a different junction type. +: N-well on substrate (NS). o: n-diffusion on substrate (nS).

Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) (configuration II):

The junction coefficient analysis continues with Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) in test configuration II. Svejk die #2 (SVJ2) in test configuration I will be considered separately. The junction coefficient plot can be seen in Figure 5.14. Note: points corresponding to N-well on substrate (NS) and n-diffusion on substrate (nS) junction types are now fits rather than pairings.

The coefficogram of Figure 5.14 reveals some interesting features. To begin with we note how the coefficients that are based on fits (N-well on substrate (NS) and n-diffusion on substrate (nS) represented by '+' and 'o' signs respectively) show much smaller differences between them than those arising from device pairings (p-diffusion on N-well (pN) represented by 'x' marks). The superposition of data from Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) thus helps confirm that the junction coefficients extracted from each junction type are unlikely to be products of mere random noise. Of course, an alternative explanation would be that we are measuring and recording the dominant sources of systematic error rather than the junction coefficients, and that these errors are junction type-specific. More data would be necessary in order to determine which effect we are observing.

Furthermore, we see that N-well on substrate (NS) devices feature only negative junction coefficients whilst the other types show negative areal and positive side-wall junction coefficients. This could be an indication that in general, pn-junctions work 'naturally' in opposition to pick-up noise (A_{p-p}) causing our estimate of electro-optical modulation (B_{p-p}) to take negative values. Side-wall junctions operate in much the same fashion, but in this technology too the presence seems to harm nearby areal component performance. N-well on substrate (NS) junctions have negative side-wall junction coefficients, but the two diffusion-based junction types show positive values instead. We expect N-well on substrate (NS) junctions to have side-wall regions that can be described as thick, vertical[19] strips that fence the N-well

[19]I.e. perpendicular to the die surface.

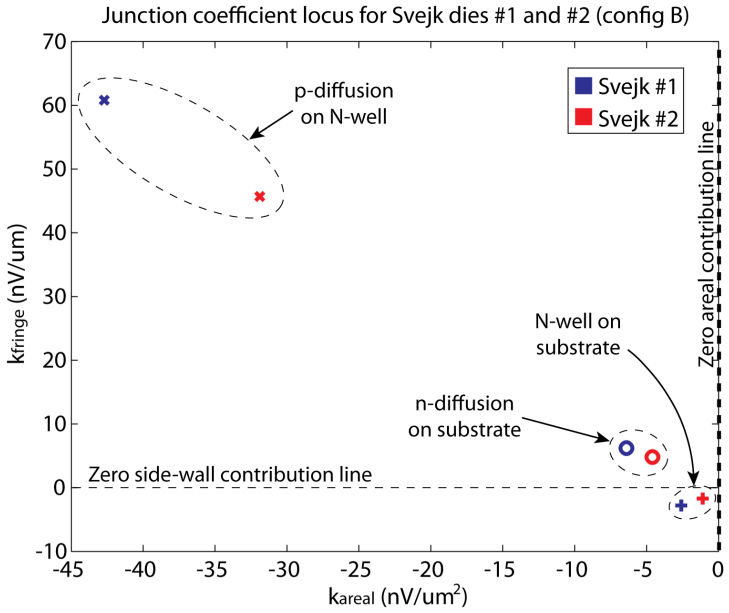

Figure 5.14: Normalised electro-optical modulation contributions of areal and side-wall junction components, pooled by junction type, for Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) in test configuration II. Each colour represents a different die. Blue: Svejk die #1 (SVJ1). Red: Svejk die #2 (SVJ2). Each symbol represents a different junction type. +: N-well on substrate (NS). o: n-diffusion on substrate (nS). x: p-diffusion on N-well (pN).

designed volume[20] whilst diffusion-based junctions are expected to have much narrower corresponding strips. This hints towards the possibility that the ratio of areal/side-wall electro-optical modulation contributions of N-well on substrate type (NS) junctions is far higher than it is for diffusion-based junctions. Should this prove to be true it constitutes grounds for expecting diffusion-based side-wall junction coefficients to be much weaker than their N-well on substrate (NS) competitors. If, on top of this, we consider the possibility that diffusion-based junctions compromise the function of large regions of what nominally constitutes areal junction territory, simply by their presence, we can explain the reason why some side-wall power coefficients can show opposite signs to those of their areal counterparts and why this should be most pronounced in diffusion-based junctions. However, as we have seen in 5.5.2.2 this is a highly speculative hypothesis and does not explain why in the case of Ninja (NIN) devices, n-diffusion on substrate (nS) junctions show positive areal junction coefficient values.

Svejk die #2 (SVJ2) (configuration I):

Junction coefficient analysis on Svejk die #2 (SVJ2) in test configuration I could be a means of testing whether with test configurations that are radically different in implementation, but very similar in principle, we can obtain at least qualitatively similar results. This analysis, however was performed in section 5.5.2.2 already. Nevertheles, for the sake of completeness the resulting coefficogram is shown in Figure 5.15.

Similar to the conclusions in section 5.5.2.2 we see a broad, qualitative similarity to corresponding coefficients extracted from both Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) in test configuration II, but still not a 'tight' match.

[20]The volume of semiconductor, characterised by a net n-type doping that corresponds to the layout shape of an N-well region constitutes its 'design volume'.

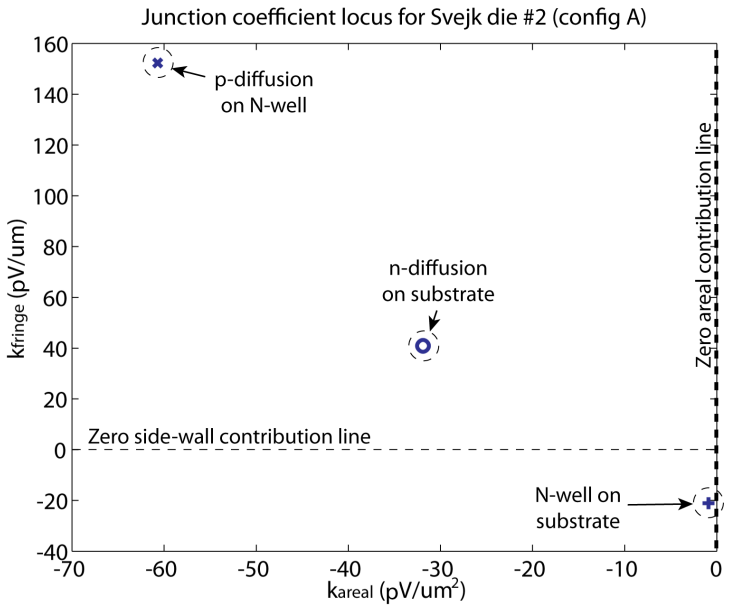

Figure 5.15: Normalised electro-optical modulation contributions of areal and side-wall junction components, pooled by junction type, (the 'junction coefficients') for Svejk die #2 (SVJ2) in test configuration I. Each symbol represents a different junction type. +: N-well on substrate (NS). o: n-diffusion on substrate (nS). x: p-diffusion on N-well (pN).

Teddy die #5 (TED5) and Teddy die #6 (TED6):

The junction coefficient scatter plot for Teddy die #5 (TED5) and Teddy die #6 (TED6) introduces an extra criterion by which we can discriminate between data points. On top of discriminating by die of provenance and junction type we can now discriminate by 'junction complexity', i.e. between what we shall call 'simple junctions' and 'complex junctions'. We shall define a junction as belonging to the 'simple' class if it is resides in a device constituted by the minimum number of doped regions necessary to give rise to that junction. Thus, an N-well on substrate (NS) junction residing in an N-well on substrate (NS) type device will belong to the simple category, as will a triple well on N-well (3N) junction in a triple well on N-well on substrate (3NS) device and an n-diffusion on triple well (n3) junction in an n-diffusion on triple well on N-well on substrate (n3NS) device. We discount cases such as the possible n-diffusion on triple well on substrate (n3S) case as we have not worked with such devices. Complex junctions will therefore be those that contain unnecessary dopant species in the immediate vicinity of the junction under consideration. N-well on substrate (NS) junctions in triple well on N-well on substrate (3NS - extra triple well) or triple well on N-well (3N) junctions in n-diffusion on triple well on N-well on substrate (n3NS - extra n-diffusion) devices are examples of complex junctions. On the basis of this categorisation, the resulting junction coefficient scatter plot will take the form of Figure 5.16.

In Figure 5.16 we have not plotted certain junction coefficients corresponding to complex junctions because they lie far outside the range within which all simple junction coefficients are clustered. Some, however, have been left for completeness. The fact that we had to leave most complex junctions out of the plot because otherwise the entire region currently shown in 5.16 would have been reduced to occupying a small spot in the 'all-inclusive' diagram (a fact easily verifiable by a quick inspection of table 5.24) testifies to the fact that the presence of extra dopants in the vicinity of a simple pn-junction under test seems to dramatically affect the performance of the junction under test. Only triple well on N-well

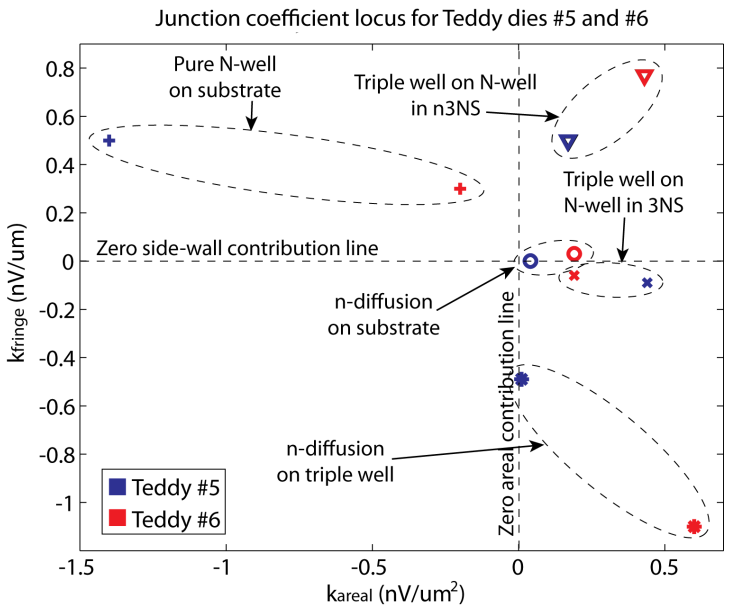

Figure 5.16: Normalised electro-optical modulation contributions of areal and side-wall junction components, pooled by junction type, (a.k.a. 'junction coefficients') for Teddy die #5 (TED5) and Teddy die #6 (TED6). Colours represent different dies. Blue: Teddy die #5 (TED5). Red: Teddy die #6 (TED6) Each symbol represents a different junction type. +: pure N-well on substrate (NS). x: triple well on N-well (3N) in triple well on N-well on substrate (3NS) o: n-diffusion on substrate (nS). *: n-diffusion on triple well (n3). ∇: triple well on N-well (3N) in n-diffusion on triple well on N-well on substrate (n3NS).

(3N) in n-diffusion on triple well on N-well on substrate (n3NS) junctions seem to have been reasonably unaffected by this phenomenon although their performance is still reasonably clearly differentiable from that of their triple well on N-well (3N) in triple well on N-well on substrate (3NS) counterparts. The reasons for this are unknown.

In terms of the junction types that have been plotted, we see a mixed picture. The differences between different junction types are comparable to their apparent spreads, particularly for 'pure' N-well on substrate (NS), n-diffusion on triple well (n3) and to a lesser extent for triple well on N-well (3N) in n-diffusion on triple well on N-well on substrate (n3NS) junctions. Triple well on N-well (3N) in triple well on N-well on substrate (3NS) and n-diffusion on substrate (nS) junctions, on the other hand, tend show smaller apparent spreads. In any case, samples from many more dies may be needed in order to determine the real spreads and either confirm or disprove the hypothesis that each junction type features different junction coefficient spreads. Thus, a lot could be potentially stated about the relations between what appears to be the 'true' junction coefficient region of each junction type, but in the absence of more measured and extracted data it would be a highly speculative undertaking. We can, however, note that the UMC13 technology provides us with a much murkier picture than IBM18, largely because it gave rise to what in comparison to other technologies are junction coefficients of very small magnitude ($< 2\,nV/\mu m^2$ and $< 1\,nV/\mu m$ compared to similar values in the 10s to 100s of normalised nV for the other technologies).

Conclusions

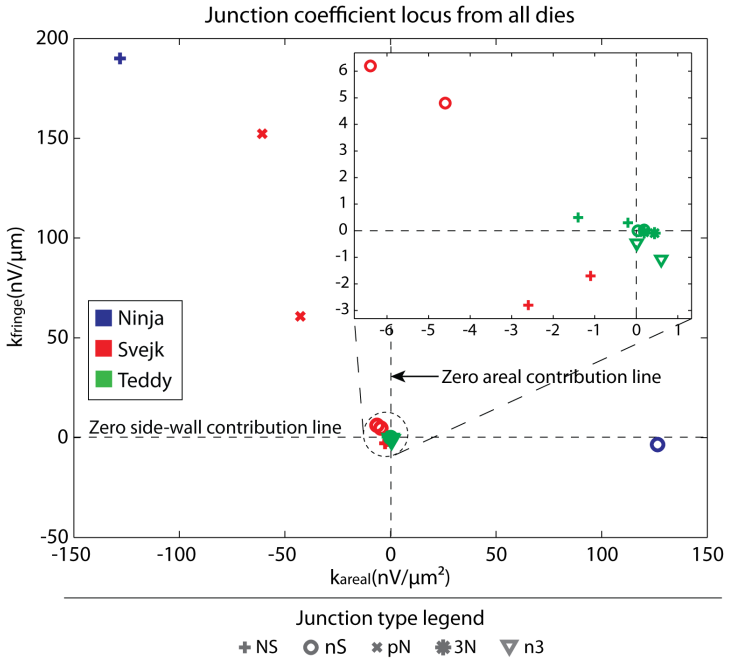

Figure 5.17: Normalised electro-optical modulation contributions of areal and side-wall junction components (a.k.a. 'junction coefficients') locus for all 'interesting' junction types from all our dies. Colours represent different die designs. Blue: Ninja. Red: Svejk. Green: Teddy. Each symbol represents a different junction type. +: pure N-well on substrate (NS). x: p-diffusion on N-well o: n-diffusion on substrate (nS). ∇: n-diffusion on triple well (n3) *: triple well on N-well (3N) in triple well on N-well on substrate (3NS).

An overview of all 'interesting' junction coefficients[21] can be seen in the scatter plot of Figure 5.17. The scatter plot makes it clear that junctions of different types do not tend to cluster together well. This indicates that junction type may not be the major determining factor when it comes to modulation performance. The technological node, however does seem to make a large difference. More measurements of more devices and more dies are required in order to draw statistically significant conclusions.

In summary, studying junction coefficients specifically in order to understand how different junction types affect modulator performance we have reached the following conclusions:

- In general we notice that junction coefficient data is insufficient to draw solid conclusions that would allow us to differentiate between junction types along clear lines. Lack of data for Ninja die #1 (NIN1) and confusing and unclear data from Teddy die #5 (TED5) and Teddy die #6 (TED6) can not be compensated for by slightly higher quality data from Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Thus all results and conclusions obtained are to a certain degree speculative estimates based on limited information.

- We have obtained junction coefficients of different signs that at least in the case of the Svejk (SVJ) design family, look reasonably repeatable. Thus, we have considered the possibility that the 'natural' behaviour of modulator devices is to work against the natural pick-up noise (measured via A_{p-p}) based on data from Svejk (SVJ) dies, but the hypothesis does not fare well in the face of

[21] I.e. coefficients extracted from pn-junctions residing in devices that have the minimum number of doped regions (p- or n-diffusion, triple- or N-well etc.) required in order to support the junction in question.

positive side-wall coefficients for Svejk (SVJ) devices, generally positive areal junction coefficients for the Teddy (TED) design family and generally a wide and inconsistent distribution of junction coefficient signs when taken across all technological nodes (also see 5.5.2.4 for a few more points of information on cross-technology modulator behaviour).

- We have seen qualitative agreement, but large quantitative discrepancies (i.e. the trends are similar, but the numbers do not match) between the relations prevailing amongst different junction types on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) in both test configurations. This improves the chances that our hypothesised trend, whereby as we move from N-well on substrate (NS), to n-diffusion on substrate (nS) and then p-diffusion on N-well (pN) type junctions in technology IBM18, areal junction coefficients tend to decrease whilst side-wall coefficients tend to increase, is indeed a valid observation.

- The rather confusing data from the Teddy (TED) design family has offered little insight into how different junction types behave in relation to one another. This is because the amount of available data is simply insufficient to allow us to discern any solid pattern in what appears to be clusters of junction coefficients characterised by large spreads and small inter-cluster distances. A side-effect of this result is that data from Ninja die #1 (NIN1), where only a single data point is provided for each junction type, can now under no circumstances be assumed to represent 'reasonable' estimates of the true junction coefficient values. Only if both Teddy (TED) and Svejk (SVJ) design families showed strong inter-die consistency could we have potentially treated Ninja (NIN) data with the assumption that it too would show the same consistency if compared to peer, AMS35 dies.

5.5.2.4 Effect of technology on performance

In order to search for trends in modulator performance that manifest themselves across different CMOS manufacturing technologies it would make sense to examine junction coefficients extracted from sets of homotype junctions hosted on homotype devices residing on the same die and compare them with similar sets from other test dies that we have fabricated. The result will be a junction coefficient scatter plot where each individual die contributes one data point per junction type. Examining available junction types on each die design family we notice that all three design families can offer information on both N-well on substrate (NS) and n-diffusion on substrate (nS) type junctions. Finally, only tests ran under configuration II will be compared. The resulting coefficogram is shown in Figure 5.18.

The junction coefficient scatter plot shows us that in terms of N-well on substrate (NS) junctions we have relatively small differences between the IBM18 and UMC13 technologies but the difference is dramatic compared to the AMS35 result. Why AMS35 features almost two orders of magnitude stronger (and therefore better) areal and side-wall junction coefficients is not easy to understand, particularly under the expectation that higher doping concentrations (and all that that implies) in technologies of finer feature sizes will lead to improved modulation performance.

Note: the AMS35 die was the thickest, but it is also substrate doping concentration that determines beam 'attrition' through the substrate. As such, it is not trivial to determine exactly how attrition through the substrate affects results. Nevertheless we can state that due to generally low doping con-

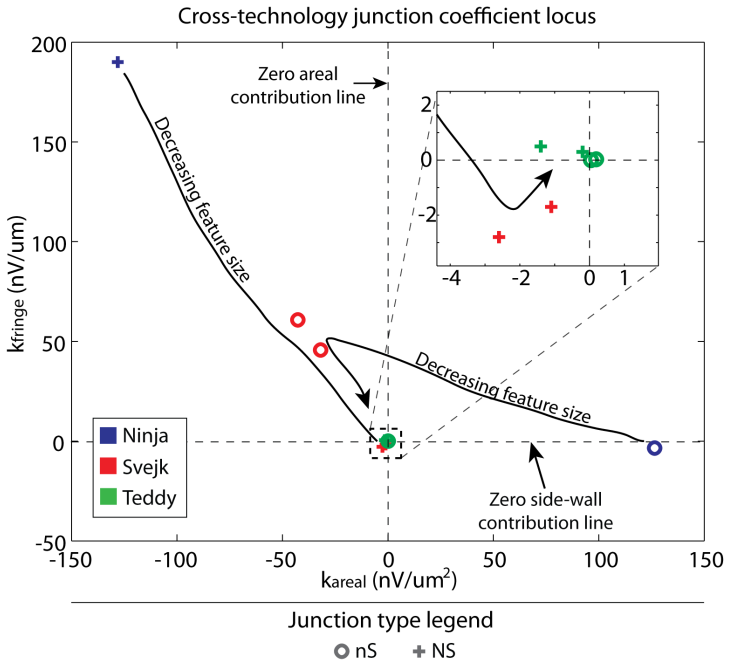

Figure 5.18: Normalised electro-optical modulation contributions of areal and side-wall junction components (a.k.a. 'junction coefficients') scatter plot summarising junction coefficients for N-well on substrate (NS) and n-diffusion on substrate (nS) type junctions across all technologies used throughout this project. Each colour represents a technology. Blue: AMS35 (NINJA). Red: IBM18 (SVEJK). Green: UMC13 (TEDDY). Each symbol represents a junction type. +: N-well on substrate (NS). o: n-diffusion on substrate (nS). Black arrows show trends as the technology feature size decreases. The inset is a magnified version of the area inside the dashed box. Surprisingly, lower feature sizes show higher and therefore better junction coefficients.

centrations in CMOS substrates we can expect that attrition though hundreds of microns of substrate is of similar significance to the attrition through the much thinner, but also much more heavily doped diffusion layer.

n-diffusion on substrate (nS) junctions show very clear differences between technologies, although it is difficult to understand why as feature sizes shrink the magnitude of the obtained junction coefficients keeps tending towards zero. Simultaneously there is no obvious reason why in AMS35 and UMC13 the side-wall coefficient should be so insignificant compared to its areal counter part whilst in IBM18 they are both of the same order. Assuming the gathered coefficients are sufficiently accurate representations of modulation capability we can only speculate that the specifics of how each junction is manufactured in each technology play a more significant role that merely the feature size attribute. This is further corroborated by the fact that it is not only for n-diffusion on substrate (nS) junctions, but also for N-well on substrate (NS) junctions that the 'trend arrows' do not follow straight lines or even curves that tend towards some limit or more generally change smoothly and according to simple rules. Instead, they both feature acute 'bends'.

5.5.3 Results within the context of a data transmission system

Analysing modulator performance in qualitative terms, i.e. understanding how each square micron of areal junction and each micron of side-wall junction contributes to overall light modulation, is a useful undertaking that provides a general understanding of pn-junction behaviour and highlights the relative strengths and weaknesses of each junction type and technology. However, the practicalities of layout will impose additional constraints on our modulator designs. It is most enlightening to define metrics such as areal and side-wall junction densities and understand the limits within which these values move in real designs. Areal junction density will refer to the coverage of a device 'footprint' area by a certain doped region and will be measured in square microns of doped region per square micron of device footprint (i.e. dimensionless). Similarly side-wall junction density will refer to the length of side-wall junction components within a square micron of device footprint area and will be measured in microns of junction side-wall per square micron of device footprint (i.e. S.I. units of 1/m).

To illustrate the importance of this concept we need simply consider that diffusion-based junctions can be laid out with much smaller minimum feature sizes than well-based structures. Simultaneously, common rules in CMOS technologies state that no point of any doped region should find itself at an xy-distance (i.e. distance along a plane parallel to the surface of the die ignoring the 'depth' coordinate) of more than a certain limit value R from a substrate bias contact. These rules tend to cover all doped regions because their objective is that substrate regions lying underneath doped regions need to make good ohmic contact to a GND bias point in order to avoid taking voltage values substantially different from GND (although, of course, without performing an exhaustive search of all design rules for every available manufacturing technology we cannot be absolutely certain that this tenet holds true).

From these two considerations we can infer the following: a) in the absence of any other rules the areal junction density for all doped species will tend towards unity as R tends to infinity. b) areal junction density will suffer because of design rules regulating how big the gaps between doped regions need to be since we will always need to 'punch holes' into our extensive doped regions in order to position substrate bias contacts (themselves containing doped regions) at strategically chosen points. Side-wall

junction densities, on the other hand, will be higher with smaller allowable layout feature sizes, which gives diffusion-based junctions the advantage.

Thus we have illustrated how design rules will impact practical modulator design; a process that will not only have to take into consideration the relative strengths and weaknesses of each junction type in terms of junction coefficients, but also their relative strengths and weaknesses in terms of allowable junction densities and furthermore, potentially also a host of other factors, such as their versatility to fit in footprints of unorthodox shapes etc.

Finally, we must note that all the obtained data refers to experiments with a specific light source. An analysis of how illumination spectral content can influence measured electro-optical modulation is shown in section 3.2.4 on page 86 where we've shown that fractional modulation depth is dependent on light source emission spectrum, photodetector spectral responsivity and transmittance as a function of wavelength (via the absorption coefficients of free electrons and free holes).

Results normalised to device footprint area

Normalising absolute modulator performance (measured by B_{p-p}) to device footprint is an exercise that may seem futile given the variety of possible layout designs available to the engineer. Moreover, there is no guarantee that the designs used for this project will represent useful models as would be, for example, 'corner cases' where we compare the design with the maximum absolute density of side-wall junctions to the design with the maximum possible areal density. Nevertheless, if there are gross differences in junction type performance we can expect that to at lest partially show. This in itself is useful because it will provide insight into the modulator behaviour of certain junction types that have no pairing partners in the library of our device designs. Moreover, we can expect that if for certain junction types one component is far more important than the other (areal vs. side-wall), this is also very likely to show given our modulator design choices.

For the above-stated reasons we shall provide tables that summarise the normalised performance of each junction alongside junction density data. The footprint sizes and B_{p-p} values for each device will also be given for quick reference. Moreover, for the Svejk (SVJ) and Teddy (TED) design families where more than one die is available we shall pool their B_{p-p} data in order to obtain what should be a slightly better estimate of their 'true performance' (mean of distribution).

We begin the analysis by examining the footprint-normalised modulator performance table for Ninja die #1 (NIN1) (table 5.26). Table data can be visualised in Figure 5.19.

Table 5.26 and the associated Figure 5.19 show us an apparently clear picture: As expected, diffusion junctions lead the table whilst N-well on substrate (NS) structures trail behind. However, despite the orderly arrangement of Ninja (NIN) devices by junction type the issue that remains is that Ninja design, dev. #5 (NIN-5) and Ninja design, dev. #6 (NIN-6) (both N-well on substrate (NS) type) generate negative estimated electro-optical modulation (B_{p-p}) values in contradiction to all other devices, including some of N-well on substrate (NS) type. Furthermore, the magnitudes of the generated, negative B_{p-p} values are larger than some positive B_{p-p} magnitudes generated by non-N-well on substrate (non-NS) devices. The significance of this is unclear.

Interesting, possible, trends seem to appear in both the figures for n-diffusion on substrate (nS) and

Table 5.26: Footprint area-normalised estimated electro-optical modulation levels (B_{p-p} values) for each device on Ninja die #1 (NIN1) (P/F) sorted in descending order.

MODULATION DENSITY: NIN						
DEV	**TYPE**	$\overline{B_{p-p}}$	**Footprint**	**P/F**	D_A*	D_P*
		mV	μm^2	mV/mm^2	$\mu m^2/\mu m^2$	$\mu m/\mu m^2$
4	nS	26.1	229441	113.75	0.90	0.05
9	nS	8.9	229441	38.79	0.34	1.33
2	pN	6.1	229441	26.59	0.37	0.05
2	NS	2.3	229441	10.02	0.12	0.66
1	NS	2.2	229441	9.59	0.33	0.27
6	NS	-4.4	229441	-19.18	0.58	0.16
5	NS	-10.2	229441	-44.46	0.58	0.16

*Areal (D_A) and side-wall (D_P) junction densities respectively.

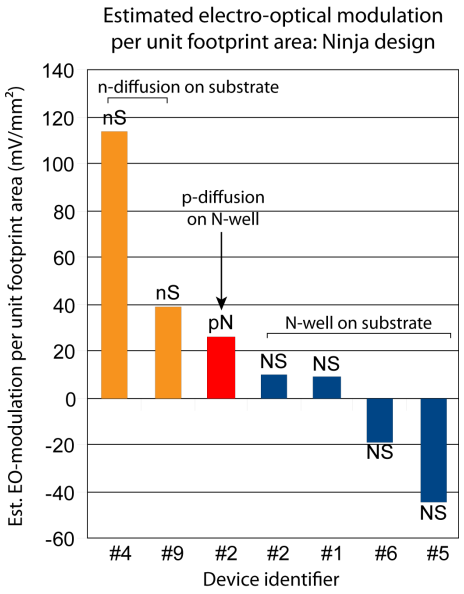

Figure 5.19: Estimated electro-optical modulation per unit footprint area for each device on Ninja die #1 (NIN1). Letters above/below each bar denote the junction of each device to which each bar refers.

for N-well on substrate (NS) junctions. For n-diffusion on substrate (nS) junctions we observe that the areal-dominated junction is overpoweringly outperforming its more side-wall dominated counterpart. Whether this observation does indeed represent a solid trend can only be confirmed or disproved by testing devices with different junction densities lying between the extremes given by Ninja design, dev. #4 (NIN-4) and Ninja design, dev. #9 (NIN-9). Meanwhile, N-well on substrate (NS) junctions show steadily lower normalised B_{p-p} values as they become more areal-dominated. This, of course, could represent either an effect related entirely to the areal element, or entirely to the side-wall element or both, depending on how much of the performance they each contribute in absolute terms. Interestingly, Ninja design, dev. #6 (NIN-6) underperforms Ninja design, dev. #5 (NIN-5) despite lacking passivation; a layer expected to perhaps weaken the illuminating beam before entering the die.

Next, we consider the footprint-normalised performance of the Svejk (SVJ) design family. Results are summarised in table 5.27 and illustrated in Figure 5.20.

Table 5.27: Footprint area-normalised estimated electro-optical modulation levels (B_{p-p} values) for each device on Svejk (SVJ) (P/F). B_{p-p} results between dies are averaged to yield a better estimate. Device # 12, N-well on substrate (SVJ-12-NS) has been removed as it was a clear outlier (reason unknown).

MODULATION DENSITY: SVJ						
DEV	**TYPE**	$\overline{B_{p-p}}$	**Footprint**	**P/F**	D_A*	D_P*
		mV	μm^2	mV/mm^2	$\mu m^2/\mu m^2$	$\mu m/\mu m^2$
4	pNS	-46.4	90000	-515.56	N/A	N/A
5	nS	-46.7	90000	-518.33	0.43	0.28
3	pNS	-47.2	90000	-524.44	N/A	N/A
1	nS	-49.55	90000	-550.56	0.13	0.00
2	nS	-52.55	90000	-583.89	0.13	0.00
6	nS	-55.3	90000	-614.44	0.24	0.00
12	3N	-29.3	40000	-732.50	0.35	0.58
12	3NS	-30.2	40000	-755.00	N/A	N/A
3	NS	-69.8	90000	-775.56	0.35	0.00
4	NS	-71.1	90000	-790.00	0.08	0.00
7	NS	-38.45	40000	-961.25	0.18	0.56
10	NS	-40.85	40000	-1021.25	0.27	0.58
11	NS	-43.05	40000	-1076.25	0.61	0.24
9	NS	-48	40000	-1200.00	0.40	0.41
8	NS	-51.95	40000	-1298.75	0.27	0.58
3	pN	-140.65	90000	-1562.78	0.34	0.19
4	pN	-144.1	90000	-1601.11	0.08	0.00

*Areal (D_A) and side-wall (D_P) junction densities respectively.

The Svejk (SVJ) design family offers us a much more mixed picture than Ninja (NIN) did, but there are some recognisable trends still present. A block of n-diffusion on substrate (nS) and p-diffusion on N-well on substrate (pNS) devices featuring similar P/F values (within $10\,mV/mm^2$ of each other) tops the chart. A solid block of n-diffusion on substrate (nS) devices follows, itself followed by the lone triple well on N-well (3N) and triple well on N-well on substrate (3NS) devices. A solid block of N-well on substrate (NS) devices then takes up most of the remainder of the table with the two p-diffusion on N-well (pN) devices constituting the bottom.

By examining the data related to n-diffusion on substrate (nS) junctions in table 5.27 we can im-

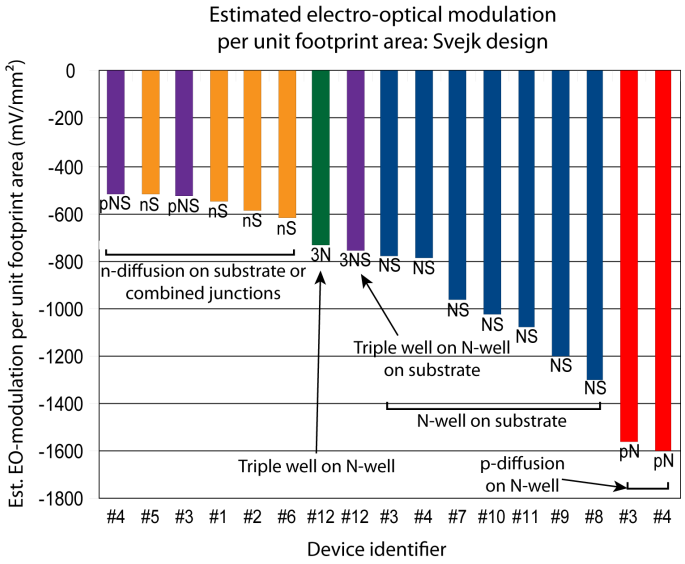

Figure 5.20: Estimated electro-optical modulation per unit footprint area for each device on Svejk dies #1 and #2 (SVJ1,2). Letters above/below each bar denote the junction of each device to which each bar refers. Device # 12, N-well on substrate (SVJ-12-NS) has been removed as it was a clear outlier (reason unknown).

mediately understand why the areal junction coefficient is negative whilst the side-wall one is positive. This is merely confirmation of our junction coefficient analysis and points towards the same conclusion: pn-junctions that will yield high magnitudes of B_{p-p} are likely to feature large exposed areas and keep their perimeters hidden behind metal masks.

The triple well on N-well (3N) junction seems to perform somewhere in the region between the n-diffusion on substrate (nS) and N-well on substrate (NS) types. Of course the absolute footprint-normalised B_{p-p} values will vary with junction densities, and given the small differences between the triple well on N-well (3N) value and those of its nearest n-diffusion on substrate (nS) and N-well on substrate (NS) neighbours, it is not inconceivable that junctions designed with an eye towards maximising the preponderance of the triple well on N-well (3N) junction type within the device may outperform some of the currently listed N-well on substrate (NS) devices (if testing for high normalised B_{p-p} magnitude).

The triple well on N-well on substrate (3NS) device configuration may have performed so closely to its triple well on N-well (3N) component alone possibly because the performance is dominated by the triple well on N-well component (3N) when Svejk design, device no .12 (SVJ-12) is operated in triple-well on N-well on substrate (3NS) mode (there is no obvious reason why this would hold true from a theoretical point of view). Alternatively the cause could be that changes in the behaviours of both triple well on N-well (3N) and N-well on substrate (NS) components simply cancel each other out when switching from triple well on N-well (3N) to triple well on N-well on substrate (3NS) operation. Once again, that is a matter that requires further study, including detailed theoretical analysis.

N-well on substrate (NS) devices, though mostly clustering in one solid block, do offer some exhibitions of strange behaviour. For example if we accept that a device with no exposed area and no exposed perimeter should be completely unresponsive to light (i.e. modulation equals zero), then it is surprising to see that Ninja design, die #4 (NIN-4) causes more modulation in absolute terms than Ninja design,

die #3 (NIN-3). The B_{p-p} value for Ninja design, dev. #3 is lower in magnitude than that for Ninja design, dev. #4. This might be explained by, for example, assuming that there will be a certain baseline change in B_{p-p} when switching the illumination on and off[22]. If this presumed baseline change in B_{p-p} is extracted from the obtained values in order to improve the accuracy of the information contained in the B_{p-p} metric, we may end up changing the signs of the B_{p-p} values for both devices. This reverses the magnitude inequality relationship between them and we find out that Ninja design, dev. #4 (NIN-4) is indeed a weaker electro-optical modulator than Ninja design, dev. #3 (NIN-3). Such distortion of B_{p-p} values could be explained by the phenomenon examined in the appendix (B.2.4), for example.

p-diffusion on N-well (pN) devices show themselves to be the bottom performers according to the obtained data. Once again, the fact that a much smaller p-diffusion on N-well (pN) type junction has created a larger magnitude of B_{p-p} seems to hint in the same direction as the discrepancies between N-well on substrate (NS) data discussed above, although care must be taken to point out that the said p-diffusion on N-well (pN) devices may be influenced by the specific way in which they have been laid out on top of their N-well 'host'.

Finally, we can examine the normalised estimated electro-optical modulation level (B_{p-p} value) table for the Teddy (TED) design family. This is shown in table 5.28 and Figure 5.21.

Table 5.28: Footprint area-normalised estimated electro-optical modulation levels (B_{p-p} values) for each device on Teddy (TED) (P/F). B_{p-p} results between dies are averaged to yield a better estimate.

MODULATION DENSITY: TED						
DEV	**TYPE**	$\overline{B_{p-p}}$	**Footprint**	**P/F**	D_A*	D_P*
		mV	μm^2	mV/mm^2	$\mu m^2/\mu m^2$	$\mu m/\mu m^2$
9	3N	13.15	89401	147.09	0.53	0.27
8	3N	12	89401	134.23	0.52	0.27
7	3N	8.85	89401	98.99	0.70	0.09
7	n3	8.8	89401	98.43	0.59	0.09
3	nS	22.8	245025	93.05	0.81	0.09
1	NS	21.1	245025	86.11	0.14	0.54
2	nS	20.85	245025	85.09	0.52	1.82
7	NS	5.5	89401	61.52	0.81	0.09
5	3N	4.3	245025	17.55	0.18	0.53
4	NS	-13.6	245025	-55.50	0.89	0.05
8	n3	-7.4	89401	-82.77	0.22	0.18
5	NS	-41.1	245025	-167.74	0.75	0.07
8	NS	-16.9	89401	-189.04	0.82	0.09
9	NS	-26.05	89401	-291.38	0.84	0.09

*Areal (D_A) and side-wall (D_P) junction densities respectively.

The Teddy (TED) design family is also providing us with a mixed picture. The only consistent feature seems to be that triple well on N-well (3N) junctions top the table. Other than that, there don't seem to be very large differences between the other junction types, thus resulting into junctions of many types being mixed together in the ranking table.

Triple well on N-well (3N) junctions behave reasonably understandably with the top two entries in

[22]I.e. even without modulation B_{p-p} will change with changing illumination, as would be the case if we were illuminating the photodetector with light bypassing the modulator structure.

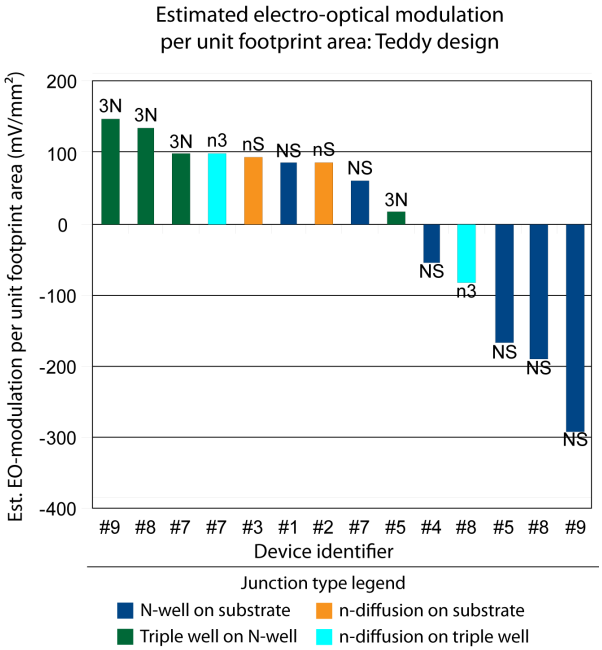

Figure 5.21: Estimated electro-optical modulation per unit footprint area for each device on Teddy dies #5 and #6 (TED5,6). Letters above/below each bar denote the junction of each device to which each bar refers.

the table being of similar junction densities (both areal and side-wall) and behaving rather similarly. Results seem to suggest that in order to improve performance on these designs particular care must be given to maximising the side-wall component as a significant drop in side-wall density and an equally significant concomitant increase in areal coverage seem to damage overall performance.

n-diffusion on substrate (nS) junctions also show a certain degree of consistency whereby it would appear that the areal component is the most important to modulator performance. Of course, in the absence of any more data from devices with different combinations of areal and side-wall densities, this is hard to confirm or disprove.

n-diffusion on triple well (n3) and N-well on substrate (NS) type junctions show disconcerting results including devices that show opposite normalised B_{p-p} signs. Particularly disturbing is the comparison between Teddy design, dev. #7 (TED-7) and Teddy design, dev. #8 (TED-8) where despite similar structures the normalised B_{p-p} values have both different signs and magnitudes. Why this should occur is not clear.

Finally, the triple well on N-well (3N) device that has no counterpart available for pairing, sits closest to zero, in a zone where it is outperformed by diffusion-based junctions, but tends to outperform N-well on substrate (NS) competitors in absolute value (but not magnitude).

Thus, we can see that in terms of modulation the lines between different junction types are not drawn clearly enough to allow us to make solid conclusions without placing careful thought and consideration into the geometric specifics that characterise each designed device. In other words, if an engineer is to design an optical modulator device, design choices relating to device geometry will be as important as choices relating to junction type. Only if the limits imposed by each junction type in terms of maximum modulator performance were significantly different, would engineers be compelled to use certain junctions

much more often than others according to the task requirements at hand. Therefore, given that analysing devices by considering performance normalised to footprint area did not give clear-cut results in most of the design families, this subject needs to be the target of a much more detailed investigation. This view is also enforced by the fact that no clear and unequivocal indications could be found that areal components should be preferred over side-wall counterparts or vice versa for each junction type. The very fact that we have obtained mutually contradictory results places the results held in all our result tables under the shadow of doubt.

Final note: much of the analysis conducted here mirrors findings during the junction coefficient analysis, however it was deemed important to consider the matter from the point of view of an engineering design involving layout constraints. Moreover, the attempt to understand, or at least estimate, how the performance of devices that have no available pairing partners fits into the bigger picture of modulators was deemed necessary.

5.6 Additional considerations

In this section we shall discuss all phenomena that have not been taken into consideration during the interpretation of our measured results. These can be subdivided into challenges to our key assumptions, i.e. potential deviations from the model implicitly used to interpret our results, and other sources of error that are deemed less likely to severely interfere with results. Discussing these non-idealities can then be broken down into an assessment of the severity, or impact, on measured results followed by an attempt to understand what causes the non-ideality to manifest itself and finally a summary of actions taken to address the situation.

Furthermore, in an attempt to assess the extent of result distortion caused by making our key assumptions we shall revisit the extra information present in the round of experiments on Svejk die #2 (SVJ2) in test configuration I. The methodology used, aims of the experiment and obtained results will also be included in this section of the chapter. Finally, the section will conclude with a small collection of conclusions that are believed to help design better electro-optical modulators in CMOS Silicon.

5.6.1 Challenging the key assumptions

Throughout this chapter we have presented and interpreted the results arising from measurements performed on each designed pn-junction on the basis of: a) information on the test set-up, b) theoretical analysis from previous chapters, c) trends inherent in the data and d) certain key and certain secondary assumptions. Within our assumptions lurk factors that may influence results that have implicitly not been tackled in section 5.5. In this part of the thesis we shall concentrate on challenging our key assumptions only. Our key assumptions can be seen in the 'Key assumptions' section on page 187 for quick reference. In this section, we shall challenge each one of our key assumptions in turn.

Inter-trial variation

Our first assumption was that A_{p-p} in the dark should remain for all intents and purposes unaltered between trials performed on the same device. Any fluctuation of A_{p-p} dark between trials will introduce

uncertainty into the system that has not been accounted for during result interpretation. Examination of this effect in practice is needed.

Effects of inter-trial variation: A very important observation to be made is that both 'dark' and 'illuminated minus dark' results are corrupted by factors that cause inter-trial variation (i.e. performance variation between successive test runs on the same device). It is important to attempt an assessment of the significance of this factor.

The full details of the corresponding supporting experiment are described in the appendix (section B.2.7). The experiment gives an example of how performance changes between trials of the same devices. However, given the fact that the changes noted are over the span of several days and include inconsistent alignment during successive trials we can expect that the trial to trial variation will be considerably smaller for successive illuminated-dark trial pairs taken for each device. Illuminated and dark trials were taken in immediate succession according to our test protocols and, crucially, without touching any part of the set-up save for the light source switch in between trials.

On the other hand, despite the fact that trial to trial variation was, perhaps, kept much better under control by use of the aforementioned test protocol, the difference between A_{p-p} values under illumination and in the dark always tended to be rather small compared to the baseline pick-up level (some inter-trial A_{p-p} variations are in the 10% region whilst B_{p-p} values may sometimes struggle to reach 1% of A_{p-p}). This means that when extracting 'illuminated minus dark' B_{p-p} values inter-trial variation corrupts their magnitudes quite severely in fractional terms.

Transferring this observation to trials between homologous devices residing on different dies we can state that the key B_{p-p} parameter will yield different values both because of inherent, Silicon differences between the devices that generate those results ('functional' differences) and because of all the factors that cause inter-trial variation (light source degradation, temperature variation, inconsistent alignment etc.). For that reason the inter-die differences shown in our result tables (5.5.2.1) are likely to be overestimates of true 'silicon-only' inter-die variation.

Note: Because of the not entirely random nature of these additional sources of uncertainty it is very difficult to quantify these effects or analytically tackle them.

Nature of inter-trial variation: One factor that causes inter-trial variation concerns inconsistencies in the set-up. Given the fact that the PCB used for each pair of dies was common amongst them and that the standard LCC package used to house each die was mounted into a standard LCC holder it is likely that if any systematic errors can be attributed to the set-up, then they can only be attributed to minute variations in the positioning of the mounted die within the LCC package. We shall call errors stemming from this situation 'mounting errors'.

Other factors would include temperature variations that cause the light source to behave erratically and failure to align the same device in exactly the same way with the light beam path between trials. Temperature variation is random and slow-changing. Misalignment is in principle random, but depends on a number of factors including the mechanical properties of the micropositioner.

Counteracting strategy: The limitations arising from uncertainties in alignment can only be limited by ensuring that the test set-up is fastened as securely as possible to the micropositioner. Only if this fastening is perfect can the alignment error be reduced to the inherent errors present in the microposi-

tioner itself. On the optical side, the collimator and optic fibre remained untouched by human hands to the greatest extent possible in order to minimise potential changes in the direction and target location of our light beam.

Presence of other light-dependent phenomena

The last two key assumptions, #2 and #3 are closely intertwined. The last assumption, in fact is a special case of assumption #2. Therefore we shall tackle both in this part of the report.

Determining why the system as a whole behaves so differently under illumination is not a trivial task. Compared to the dark case we have optical effects that affect the photodetector, optical effects that affect the modulator devices (excluding electro-optical modulation) and electro-optical modulation; our target phenomenon. All these effects will change according to the die that is inserted between collimator output and photodetector. Finally, we have non-optical effects that can compound optical effects and increase result distortions.

Non-optical effects: In this domain we can include the effects of amplifier non-linearities such as common-mode and input signal amplitude gain dependence and distortion. Sadly the manufacturer does not state the performance of the amplifier to distortion and we have not characterised it ourselves either, instead implicitly trusting the manufacturer. Nevertheless, the weak nature of our results imply that amplifier non-linearities may not be ignorable.

The possible effects can be summarised thus: Common mode gain dependence means that B_{p-p} depends on the magnitude of the DC current arriving from the photodetector. This means that B_{p-p} represents at least the modulation and the amplifier common mode gain dependence. Next, input signal amplitude gain dependence and any other distortionary effects in the amplifier will cause the input to the lock-in amplifier to be characterised by a broad peak. This may affect the measured value of A_{p-p} that relies on locking to 'near but not exactly at' the $8\,kHz$ mark (see section 5.4.6.).

Optical effects affecting the photodetector: We have considered one optical effect that can potentially affect the photodetector. The mechanics of the said phenomenon are discussed in more detail in section B.2.4 of the appendix, as are its effects. However, we shall discuss the subject here briefly in order to position it more into context and add some notes. Moreover, an entire subsection in this section will be devoted to special testing performed in order to try and determine the effects of this phenomenon (section 5.6.2).

Our photodetector will behave under NIR illumination (approx. $800 - 1800\,nm$ wavelength) in a fashion similar to how Silicon photodetectors behave under visible light: it will create a photocurrent. This photocurrent then, can change the bias point of the photodetector and thus alter small signal impedance, thus affecting sensitivity to extraneous stimuli such as pick-up.

In terms of strategies employed to counter-act the phenomenon, we have not employed any in our main experiments, largely because this phenomenon was not considered until a very late stage of the project.

Note: Because of the absence of any light filtering throughout our optical modulation experiments much light that belongs both to the absorption band of Silicon and Germanium (i.e. in the approx.

$800 - 1100\,nm$ wavelength band - the 'overlap' band) can conceptually reach the photodetector. This means that the effects of illumination on the photodetector are likely to be substantially different between different modulator dies if the latter offer different overall optical resistance to light in the overlap band and still allow a substantial amount of it to reach the photodetector. We do not know how much light in the overlap band does reach the photodetector after having crossed through the Silicon modulator dies, but it is certain that this effect -if present- could have been easily largely eliminated by using long-pass filters that cut off wavelengths above the absorption band of Silicon. This was not implemented, however, in an attempt to maximise the amount of light being modulated by our devices. In retrospect, that decision was mistaken.

Note: The effect of the illumination upon the photodetector will still be die-specific even if we eliminate all light lying within the overlap band. This occurs because light outside the overlap band will still be subjected to free-carrier absorption and different dies will likely offer different total FCA-based optical resistance to the incoming light. However, we expect these effects to be significantly smaller given the weakness of the free carrier absorption phenomenon; a weakness that ensures that the vast majority of light can pass through an entire CMOS die largely unabated (see chapter 4 for a plot of change in absorption coefficient as a function of free electron and free hole concentration). Nevertheless, significant differences in the degree to which different dies are covered by metallisation could still conceivably cause the effect to be significantly die-specific.

Optical effects affecting the modulator: Light within the absorption band of Silicon that falls upon the modulator die will increase the rate of free carrier generation within the semiconductor. This may upset the natural balance between generation and recombination and lead to the presence of more free carriers per unit volume. In practical terms, this means that once the set-up is illuminated the modulator die reaches an equilibrium point that is different from that in the dark from an electrical point of view (specifically: free carrier distribution and thus also semiconductor conductivity distribution). This immediately points towards the possibility that pick-up generation will be affected by the mere act of illuminating the die with EM radiation within its absorption band. In terms of our results, it means that the A_{p-p} value in the dark and under illumination will be different partly because of pick-up generation differences between illuminated and dark states. Of course, this also changes the amount of free-carrier losses through the Silicon, further adding to the complexity of the problem.

Whether such effect will indeed take place or not requires further study. Specifically, we would need to calculate the free carrier concentration at equilibrium for the dark and illuminated cases respectively, given certain assumptions about the nature of the dominant recombination mechanisms taking effect in our modulator dies. Once again, this effect, if present, can be eliminated by filtering all light that falls within the absorption band of the modulator host die. Discovering the issue late in the project precluded this option from being exercised, however.

Result distortion from other electro-optical phenomena - summary: In conclusion, we have determined that from a purely theoretical point of view we can expect the following interference with our experiments from optical phenomena other than electro-optical modulation:

1. The photodetector will likely experience a change in its small-signal impedance as a result of the received light. Thus, between measurements in the dark and under illumination respectively, we will experience a different photodetector sensitivity to pick-up. This can be combated by attempts

to minimise pick-up and maintain the bias voltage across the photodetector as close to constant as possible with electronic means.

2. If the illumination contains a significant component that lies in the overlap wavelength absorption band where both modulator host die and photodetector are sensitive[23], then it is possible that different dies will offer different optical resistances to the incoming light and thus let different amounts of light within the overlap band reach the photodetector. This will make the phenomenon described in the point above die-specific and perhaps even device-specific. This can be combated by filtering light in the troublesome wavelength band. This phenomenon will also apply for light outside the overlap band, but to a heavily reduced extent only.

3. Under the conditions described in the point above, the modulator host die may experience a change in the way it generates the pick-up signal as a function of illumination due to an alteration in the balance of free carriers within its Silicon. This can also be combated by filtering out light whose photons are energetic enough to significantly alter the rate of free carrier generation in Silicon.

Unfortunately, quantifying the influence of each effect is not trivial and neither is separating the influences of each factor on the key A_{p-p} metrics obtained for each device. Hence attributing the F_{Ap-p} differences between dark and illuminated set-ups solely to differing modulation efficiencies between homologous devices residing on different dies is not possible without further investigation.

5.6.2 Revisiting Svejk die #2 (SVJ2) in test configuration I

Running the supporting experiment described in section B.2.4 of the appendix was the result of careful consideration of the underlying theory after an assessment that results obtained from our regular tests in both configurations were unsatisfactory and deviate from expectations to the degree that further investigation is required. It was in reaction to the results of that experiment, that a special set of experiments on Svejk die #2 (SVJ2) in test configuration I was ran. The reason for choosing Svejk die #2 (SVJ2) in test configuration I was two-fold: a) The Svejk (SVJ) design family appeared to be the most 'well-behaved'. b) Certain PCBs used in configuration II had to be destroyed in order to create configuration I, which left no alternative with regards to the test configuration used.

We shall examine the test protocol used, go over the results obtained and finally attempt to understand what impact these measurements make on our system.

Test protocol

The test protocol has been briefly described in appendix B.1.1, but we shall present the key points here for convenience and add details that help place this experiment into perspective.

The experiment was based on the concept of 'natural pick-up'. Working with natural pick-up was possible when testing Svejk die #2 (SVJ2) because the PCBs used for the Svejk (SVJ) design family,

[23] And of course the associated sensitivity of the photodetector is sufficiently high to cause intolerably significant changes in photocurrent as a function of changes in irradiance within that band.

quite unlike their counterparts allocated to the Teddy (TED) and Ninja (NIN) families, were based upon the concept of BNC ports receiving the modulator bias signal from the signal generator and then distributing it via a series of switches to various devices on each Svejk (SVJ) die. Teddy- (TED-) and Ninja- (NIN-)specific PCBs, on the other hand, generally feature BNC connections hard-wired to the LCC holder and then routed into the corresponding modulator device directly.

This special feature of the Svejk- (SVJ)-specific PCBs can be exploited in the following way: the signal generator can be allowed to feed the PCB's BNC input port, but then all the switches can remain shut (i.e. the PCB is forced into its 'neutral state'). The result is a signal that can cause significant pick-up at the photodetector. The signal biases the Svejk (SVJ) PCB tracks without actually forcing any of the devices on the Svejk (SVJ) die to start electro-optically modulating the passing light. The implication is that we can now detect changes in pick-up as a function of illumination. The presence of light should in no way affect the generation of pick-up, but will presumably affect the reception of pick-up, as discussed before. Note: PCB tracks between the switches and the chip will also introduce pick-up not covered in our 'neutral state' calculations, so this method is not perfect.

This warrants experiments similar to those carried out for testing every device for optical modulation: an alignment stage, a measurement taken in the dark and a measurement taken under illumination. It may seem superfluous to attempt and align the test system when no modulation is actually being carried out, but we wish that for each device under test the illumination going through the die and reaching the photodetector to be exactly the same in profile and alignment as when testing the modulator. This means that the fractional differences in how pick-up is received between the light-on and light-off states are likely to closely match the underlying change in A_{p-p} (ignoring the modulation component or any effects on the concentration of free-carriers in the modulator host die) between corresponding tests carried out with the modulator devices connected to the signal generator.

The above point can also be thought of thus: If electro-optical modulation and light-dependent pick-up generation capability effects are ignored and the illumination between experiments is effectively of the same magnitude, alignment and profile, then the ratio of A_{p-p} in the dark over A_{p-p} under illumination should be the same regardless of whether the pick-up being measured originates at the PCBs in the neutral state, or comes from any modulator device or combination of active modulator devices. The observed change in A_{p-p} will simply be a result of changing small-signal impedance at the photodetector and thus in no way discriminate between pick-up noise generated at different sites.

Any possible hidden assumptions implicit in this updated model have not been fully examined within this project, so the statement above comes with a word of caution. Nevertheless, if we assume that the model is sufficiently accurate as is, then we can proceed and attempt to predict how such system would behave with the presence of electro-optical modulation and light-dependent pick-up generation capability effects. Furthermore, we can find a way to extract meaningful conclusions from the obtained results. For clarity we shall describe the procedure in the form of an algorithm. For each individual device:

1. Find A_{p-p} (dark) and A_{p-p} (light) in the corresponding 'neutral state' (for differentiation we will call these $\hat{A}_{p-p}$ (dark) and (light) respectively).

2. Compute the ratio of $\hat{A}_{p-p}$ light/dark (we shall call this value 'N').

3. Carry out A_{p-p} measurements in the dark and under illumination with the modulator device

connected.

4. Compute the corresponding ratio of A_{p-p} light/dark for the experiments with the modulator connected (we shall call this value 'M').

5. Multiply the A_{p-p} (dark) obtained from the modulator test by N in order to yield a predicted value for A_{p-p} under illumination with the modulator connected (we call this $\mathring{A}_{p-p}$).

6. The obtained predicted value yields an estimate of what the measured A_{p-p} under illumination should have been in the absence of any electro-optical effects. Subtract the predicted 'ideal' $\mathring{A}_{p-p}$ from the measured A_{p-p} under illumination in order to reveal the difference C_{p-p}.

7. Use C_{p-p} as an estimate of the influence of all electro-optical effects influencing the modulator die on the system.

Of course, the fact that these steps need to be carried out with proper consideration of alignment and generally good laboratory practices goes without saying. Moreover, C_{p-p} is a difference between a measured and a predicted A_{p-p} value. This immediately implies that we consider electro-optical effects affecting the modulator die to be completely separate from pure electrical and electro-optical effects affecting the photodetector (such as the change in small-signal impedance).

Finally, when taking $\hat{A}_{p-p}$ measurements we can define standard deviation values σ that place the nominal result into perspective, just as we did for A_{p-p}.

Obtained results

The results obtained from our new round of measurements are summarised in table 5.29.

These results need to be further refined in the same way as B_{p-p} results were processed previously in this chapter. Nevertheless, a quick look at the uncertainty prevailing in our measurements, combined with the observation that more often than not C_{p-p} is going to be substantially smaller than ${B_{p-p}}$[24] shows that this effort is unlikely to provide solid results. Nevertheless, we can make two 'interesting' observations.

The first is of a technical nature: When measuring $\hat{A}_{p-p}$ for Svejk die #2, dev. #3, p-diffusion on N-well junction (SVJ2-3-pN) and Svejk die #2, dev. #12, triple-well on N-well junction (SVJ2-12-3N) we encountered unusual levels of noise, as can be evidenced by their corresponding $\sigma_{\hat{A}}$ values. The noise appeared for a certain time interval during the measurement round that yielded all $\hat{A}_{p-p}$ values in the table and then promptly disappeared as $\hat{A}_{p-p}$ for Svejk die #2, dev. #12, triple-well on N-well junction (SVJ2-12-3N) was being measured in the dark. The cause of this behaviour is still unspecified but makes the point that even in measurement rounds performed within the shortest time interval possible in order to minimise sources of uncertainty, spurious phenomena will persist.

The second observation is that it would have been interesting and perhaps more educational to have performed these tests in configuration II. The generally higher A_{p-p} values measured under configuration

[24] A quick check of this fact can be carried out by checking which value, A_{p-p} in the dark or $\mathring{A}_{p-p}$ is 'closer' to A_{p-p} under illumination. Most of the time it will indeed be $\mathring{A}_{p-p}$ that is closer.

Table 5.29: Extracting C_{p-p} from our measurements on Svejk die #2 (SVJ2) devices in test configuration I.

SVJ2 C_{p-p} RESULT SUMMARY								
DEV	**Type**	**Light**	A_{p-p}	σ_A	$\hat{A}_{p-p}$	$\sigma_{\hat{A}}$	$\mathring{A}_{p-p}$	C_{p-p}
		ON/OFF	mV	mV	mV	mV	mV	mV
1	nS	ON	16.81	0.45	12.45	0.48	17.14	-0.33
		OFF	17.42	0.42	12.65	0.39	-	-
2	nS	ON	17.86	0.36	11.91	0.38	17.88	-0.02
		OFF	18.54	0.29	12.35	0.34	-	-
3	NS	ON	14.35	0.50	12.37	0.31	14.62	-0.27
		OFF	14.97	0.46	12.67	0.28	-	-
3	pN	ON	11.21	1.22	10.28	1.57	10.67	0.54
		OFF	10.50	1.65	10.12	1.40	-	-
4	NS	ON	14.99	0.30	11.96	0.36	15.06	-0.07
		OFF	15.48	0.23	12.29	0.33	-	-
4	pN	ON	7.63	0.45	5.86	0.43	7.69	-0.06
		OFF	8.04	0.46	6.13	0.44	-	-
5	nS	ON	13.08	0.21	11.93	0.37	13.05	0.03
		OFF	13.30	0.25	12.16	0.37	-	-
6	nS	ON	12.43	0.38	11.12	0.33	12.57	-0.14
		OFF	12.84	0.25	11.36	0.32	-	-
7	NS	ON	15.19	0.29	11.39	0.36	14.99	0.2
		OFF	15.56	0.20	11.82	0.32	-	-
8	NS	ON	12.73	0.22	11.51	0.38	12.91	-0.18
		OFF	13.31	0.32	11.87	0.33	-	-
9	NS	ON	13.07	0.20	10.97	0.31	13.48	-0.41
		OFF	13.57	0.37	11.04	0.36	-	-
10	NS	ON	13.25	0.26	11.70	0.42	13.39	-0.14
		OFF	13.71	0.36	11.98	0.36	-	-
11	NS	ON	13.03	0.19	11.90	0.29	12.86	0.17
		OFF	13.18	0.25	12.20	0.33	-	-
12	NS	ON	15.70	0.21	11.91	0.39	15.49	0.21
		OFF	15.75	0.23	12.11	0.36	-	-
12	3N	ON	9.44	0.39	6.16	1.38	10.89	-1.45
		OFF	9.81	0.41	5.55	0.41	-	-

We show in each column from left to right: a) Device identifier. b) Junction type. c) Light status. d) A_{p-p} with the modulator operational. e) Variation inherent in previous A_{p-p}. f) A_{p-p} values in neutral state. g) Variation inherent in A_{p-p} measurements at neutral state. h) Predicted A_{p-p} given measurements at neutral state. i) Resulting C_{p-p} value.

II, and the presumably similarly larger values of $\hat{A}_{p-p}$ that would have resulted might have been far less susceptible to noise and thus revealed a different picture. Lack of time and the destruction of vital elements of configuration II for Svejk (SVJ), however, prevented us from attempting this.

Impact on B_{p-p}-based results

The conclusion that is drawn from these results shows that our recorded B_{p-p} values must now be considered far less likely to represent the true extent of all electro-optical phenomena influencing the modulator die, let alone electro-optical modulation on its own. Further study needs to be undertaken in order to truly understand electro-optical modulation under the operating regime specified by our experiment (i.e. transmittance mode electro-optical modulation) and either confirm our current results as valid despite the heightened uncertainty, or disprove them as the results of measuring mostly noise or spurious responses.

5.6.3 Challenging secondary assumptions

Our 'secondary' assumptions, in contrast to key assumptions, tend to be of a more technical nature. Specifically, rather than considering certain physical phenomena to be of negligible importance, our secondary assumptions consider certain technical imperfections in terms of our set-up and environmental control associated with it to be of negligible effect. Thus, in this subsection we shall examine phenomena such as light bulb degradation over time and thermal effects acting on the light source.

Light bulb degradation

Throughout the duration of this project we have made a large number of attempts to measure electro-optical modulation from our devices. This means that the light-bulb inside the illuminator was used for long periods of time. Eventually, the first bulb we used failed completely. Inspection of the light bulb revealed that a black substance had been deposited upon the surface of the light bulb, thereby obstructing the generated light. Inspection of preliminary results (not shown throughout this thesis) also revealed that the measurements in the run up to the bulb failure showed a gradual, but very significant (10-20 % of peak performance lost at least - measured by means of comparing A_{p-p} between successive rounds of measurement on the same or homologous devices) decrease in measured A_{p-p} values. This brought the issue of light bulb degradation and its seriousness to our attention.

The counter-measure to this issue was to use a brand new light-bulb and take all measurements before it could clock too many hours of operation. Thus, we made every effort to ensure that all experiments were ran within a time interval much shorter than the manufacturer's rated light-bulb life-span at full power.

Thermal effects on the light source

The illuminator we used in our experiments houses a $150\,W$ light bulb that generates all light used for testing for electro-optical modulation. This will, of course, heat up to a considerable degree. Rough observations indicate that this event does not affect light output considerably, although in retrospect, a

thorough test should have been carried out fully characterising the phenomenon. Ambient temperature changes also affect the temperature at which the light bulb reaches thermal equilibrium, but the effects of this on light output have not been studied. Instead, we have assumed that since no considerable change in illumination can be seen between the initial switch-on moment and 10-15' later.

Nevertheless, in later experiments it was decided that in order to avoid any possible effects of the thermal stabilisation process on light output experiments should start after at least 10' of thermal stabilisation time and instead of switching the light source on and off when measuring A_{p-p} values under illumination and in the dark respectively, a paper cover would block the incoming light during the 'dark' measurements instead. The paper obstacle was of sufficient thickness to block all light and the choice of paper as a material was to ensure that it would introduce no changes in the pick-up signal. Previous attempts to cover the die package with metallic obstacles proved that the metallic objects used to obstruct light do indeed affect the amplitude of the pick-up signal at the photodetector. This was proven by measuring A_{p-p} in the dark with and without the obstacle present. The differences were significant (results not shown).

5.6.3.1 'Upstream' junction biasing

In our experiments, so-called 'upstream' pn-junctions were left floating. As explained before this was a result of the test-bench design not being able to easily allow upstream junction biasing during the testing operation of a target pn-junction. As a result, the biasing across each upstream pn-junction becomes essentially unknown during measurement. The input signal and the illumination both have the capability of changing the bias voltage across floating pn-junctions.

The input signal, for example, will create a complicated perturbation of the voltage bias across the pn-junction. Ideally the junction should act primarily as a capacitor (i.e. small charge transfer between p- and n-sides) because the voltage across the junction will tend to be preserved by virtue of capacitive coupling. However, at the same time charge redistribution from regions far away from the depletion region, including perhaps the metallisation used to connect to the now floating terminal of the device, will tend to counter-balance that effect. This leads to changes in biasing across the floating junction and may create complicated patterns of pick-up as well as modulate the depletion region of the floating junction and add 'parasitic' electro-optical effects when under illumination. This affects measurements in nested junctions because the severity of this effect will depend on the type and layout specifics of the floating junctions, thus creating inconsistencies between homotype devices with different layouts.

The illumination, on the other hand, will change the operating point of the floating pn-junction by generating a photocurrent that will forward-bias the said junction. This changes the small-signal impedance, thereby likely affecting how the floating junction reacts to the input signal. B_{p-p} values will be affected as a result. To what extent pick-up generation from such effects at floating pn-junctions are significant is unknown.

The conclusion is that, unfortunately, nested pn-junctions simply suffer from too many testing issues. Even if upstream junctions were properly biased during our measurements the fact -documented earlier- remains: pn-junctions that host superfluous doped species in their immediate vicinity will be affected by those doped species on a materials level in complicated ways and thus a comparison between nested-junction devices will always pose additional challenges. Nevertheless, building a test set-up that allows

properly biasing upstream junctions as well as their downstream counterparts will be mitigate these problems to a certain extent.

5.6.3.2 Beam profile and alignment errors

The issues of a non-uniform beam profile and misalignment have already been discussed to some extent in B.2.2 and 5.4.3. However, we shall summarise the relevant information from both above-stated sources here for the sake of completeness and present in a way that stresses the possible effects of these uncertainties.

In B.2.2 where we concluded that due to the presence of only one peak in the reference diode photocurrent vs. location plot and because that peak is rather broad we can consider that once properly aligned, each device has access to a reasonable approximation of uniform illumination. The only good way of testing for this will be to use a very fine-step micropositioner and obtain detailed information on beam profile on a two-dimensional grid. Repeated measurements (if the fine pitch micropositioner has excellent bidirectional repeatability) can be averaged to yield better estimates.

In terms of alignment we have already mentioned (5.4.3) that the bidirectional repeatability of the positioner stages is in the region of 2 microns. This means that when manipulating the die via the micropositioner, we are not simply moving on a grid of predetermined, possible positions, but rather due to the bidirectional repeatability value we are moving roughly on the grid with a certain uncertainty added on top of that. Whether this uncertainty builds up over time or not is unclear. If it does, then we may have a maximum alignment error of 50 microns in either direction (x- or y-direction), or equivalently land at a distance of up to approx. 71 microns ($\sqrt{2} \cdot 50\,\mu m$) away from the illumination intensity 'peak'. Any error larger than that will be corrected by stepping the micropositioner (reminder: micropositioner step = 100 microns). Given information on the beam profile and knowledge of our device sizes, this source of uncertainty is not expected to affect results very significantly. It the uncertainty doesn't build up over time, then this issue does not manifest itself.

Under our current set-up configuration, associated limitations and the results of the experiment described in B.2.2 it was deemed that neglecting alignment inconsistencies was the most realistic approach.

5.7 Design recommendations for CMOS electro-optical modulator test-benches

In light of the results and considerations shown throughout this chapter we cannot offer solid advice on what will constitute good CMOS electro-optical modulator design. However, we can offer a few hints towards what seems to be electro-optical modulator test-bench design with increased chances of successfully isolating and measuring electro-optical modulation in CMOS structures under the transmittance mode test configuration.

- Ideally, the light source and the photodetector will be forced into alignment and locked to those relative positions. This will ensure that the same beam profile falls in principle upon the same patch of photodetector area, thereby eliminating errors associated with uneven responsivity across

different parts of the photodetector exposed surface. The modulator die alone will be placed on a mobile platform capable of positioning its payload with a high degree of accuracy and repeatability.

- The emitted light beam must be properly collimated and maintain an effectively constant beam transverse cross-sectional area from the point where it is released into the air to the point where it hits the photodetector. This will minimise any errors introduced by changing the distances between the modulator die and the fixed collimator and photodetector complex.

- Circuits must be designed on the modulator die that will generate a (preferably controllable) oscillating signal to drive the modulators with the minimum amount of charge required to service the corresponding junction capacitances. Not being forced to charge and discharge comparatively large PCB line capacitances while driving the modulator should remove a major contributor to pick-up.

  The modulator driver signal will be preferably switchable between a digital pulse and a sinusoid (to help examine 'analogue' and 'digital' operation regimes of the modulator). Note, however, that generating very pure sinusoidal signals (able to compete with those provided by a signal generator) on-chip may prove difficult to achieve. This may render successful locking to the signal frequency by means of a lock-in amplifier less likely.

- The physical distance between modulator die and photodetector will have to be kept sufficiently large in order to minimise the effects of pick-up on the photodetector.

- In experiments ran for modelling purposes, the devices should be designed to feature large basic cells with areas close to their border regions (within few minimum design widths of the nominal junction edge) being subjected to the minimum possible amount of shading. This will allow possible development of an enhanced, 'border area + core area' model of modulator performance rather than relying on nominal area and side-wall regions. We envisage that the enhanced model would consider the side-wall and adjacent areal territory[25] as a lumped, 'border' region with the remaining areal components classed as 'core area'. Sadly the majority of our devices fails both the basic cell size and the shading criteria. The 'nominal area + nominal perimeter' model, is most likely insufficiently accurate for any practical purposes.

A host of other measures could be taken to ensure consistency between measurements and eliminate various sources of error, such as ensuring that the experiment is carried out in a temperature-controlled environment. However, if the optical arrangement is done with great care and the signal driving the modulators is generated on-chip and limited to the minimum amount of charge transfer, then it is possible that the pick-up element will be reduced to a degree sufficient to allow a proper measurement of electro-optical effects occurring on the modulator die. If pick-up is effectively eliminated, then measured results would ideally reveal certain, non-zero A_{p-p} values for each device and junction under illumination and be indistinguishable from zero for all devices and junctions in the dark. Then, those A_{p-p} values will be much closer to representing electro-optical modulation on-chip (albeit indirectly).

[25] Defined as areal junction regions within some distance 'd' of the nearest side-wall edge.

5.8 Conclusions

Throughout this chapter we have shown measured results extracted from our modulator structures and interpreted them within a context created by a description of the test set-up, supporting experiments and additional considerations, including challenging our assumptions.

Processing our raw results showed that key metrics considered throughout our test procedures as crucial markers of modulator performance, including A_{p-p} and B_{p-p}, suffer from large amounts of variability that can only be sufficiently contained by means of running much larger numbers of measurements on large numbers of dies and then extracting their statistics. In the absence of such information we endeavoured to gain more qualitative insights into the behaviour of our test structures within our test set-up. Thus, using information pertaining to the specifics of our test set-up and a few strategically important supporting experiments we launched a first attempt to interpret our results while maintaining key assumptions about the underlying physical behaviour of our system.

The effort revolved around extracting so-called 'junction coefficients' that are meant to quantify the contribution to electro-optical modulation of each square micron of areal junction and each micron of side-wall junction length respectively. Results revealed large variability and generally weak patterns and trends. Inconsistencies even in the sign of the obtained junction coefficients hinted towards the possibility that side-wall and areal junction components on some occasions counteract each other in terms of electro-optical modulation. Reconsidering the validity of a rigid separation between areal and side-wall junctions offered some explanation of such phenomena and led to suggestions about improved result processing techniques.

The fact that the insight offered to us by a careful examination of our test set-up and the data provided by our supporting experiments only partially clarified the situation led us to challenge our fundamental assumptions pertaining to the nature of all phenomena that occur throughout our test chain. Particular attention was paid to electrical and optical effects affecting the modulator die and the photodetector. It was found that certain effects that were neglected by assumption could no longer be excluded as sources of severe errors. The most important such effect was found to take the form of light-dependent small-signal impedance change at the photodetector. In combination with strong pick-up from the signal generator, this effect entirely overshadowed any electro-optical modulation occurring at the modulator junction. The fact that pick-up and electro-optical signal share the same frequency renders differentiating between them extremely challenging.

What is therefore certain is that electro-optical modulation is part of our results, but what is almost equally certain is that it is not the dominant effect. In practical terms it seems that the useful information is buried in the pick-up and thus we can not take the obtained results at face value.

Nevertheless, our work has served to uncover a substantial part of the full extent of technical complications related to achieving high quality electro-optical modulation in the transmittance configuration. We have identified key factors that cannot simply be assumed to be negligible and offered advice on how some may be eliminated. By far the most important improvement that can be made to any set-up concerns creating the modulator driver signal on-chip, via specialised driver circuits that consume the minimum amount of charge in order to correctly bias the modulator structures. We are confident that, should the issues we are pointing towards be taken into account, the viability of this technology can be

proven. Furthermore, we recommend that any further tests must be carried out on highly automated test-benches with large number of available dies.

In summary, we have obtained some useful information despite all difficulties: a) insights gained via the investigation of the factors that led to the demise of our results and b) advice pertaining to the design of what stands a good chance of being an adequate test platform for measuring FCA-based amplitude modulation in CMOS modulators.

Bibliography

[1] S. M. Sze, *Physics of Semiconductor Devices*. Wiley, 2nd ed., 1981.

[2] K. Nikolic, A. Serb, and T. Constandinou, "An optical modulator in unmodified, commercially available cmos technology," *Photonics Technology Letters, IEEE*, vol. 23, no. 16, pp. 1115–1117, 2011.

[3] A. Serb, K. Nikolic, and T. Constandinou, "Feasibility of an electro-optic link for bondpad-less cmos lab-on-chips," in *Biomedical Circuits and Systems Conference (BioCAS), 2011 IEEE*, pp. 353–356, 2011.

[4] A. Hastings, *The art of analog layout*. Pearson Prentice Hall, 2006.

Chapter 6

Optically powered CMOS chips

A key function of any contactless integrated circuit is power/signal recovery. In this project the power harvesting capability was implemented optically; a well-known and widespread method that is based upon the principles that drive solar cells (presented in chapter 3). Despite the fact that all optoelectronic test devices designed throughout this project have been tailored specifically for helping understand electro-optical modulation within an engineering context, the very same devices feature characteristics that render them equally suitable for helping us to deepen our understanding of optical energy harvesting.

In this chapter we present results stemming from I-V sweeps of the various optoelectronic devices residing on Ninja (NIN), Teddy (TED) and Svejk (SVJ) and attempt to find rules that describe the behaviour of these devices in 'power recovery mode', notably the individual contributions of areal and side-wall junction components, the role of the technological node etc.

From a circuit design perspective, very rarely is the raw output voltage of an optical power recovery device sufficient to provide enough headroom for any load circuits it is meant to be operating. For that reason systems that draw power from external light sources tend to either employ series-connected photo-capturing elements or specialised power management systems that upconvert the output voltage of the photo-capturing elements to a sufficiently high level and then down-regulate it to the desired voltage headroom. In this project we have concentrated on the power management system approach and designed a power management unit that according to simulations is capable of conveying low-voltage, incoming optical power to a stable, $1.5\,V$ electrical power supply. Schematics of the system, simulated performance results and an assessment of that performance are all topics covered within this chapter.

Overall, the chapter is organised as follows: We begin by considering the implications of our electro-optical device design choices on power scavenging capability (section 6.1) before moving on to summarising the raw results from each test structure (section 6.2). Before raw results are processed in section 6.4 a section is devoted to supporting experiments that help place the aforementioned raw results better into perspective. A small section on various caveats and non-idealities that corrupt the reliability of our measurements is presented in section 6.6. Attention then shifts to the design (section 6.7) and operation of the power management unit (PMU) (section 6.8). Within these sections we review the part of the design cycle of the PMU system that lies between the design choices and the simulated results. Finally a summary of the chapter closes the power harvesting section of this thesis (section 6.9).

6.1 Photodiode design for power scavenging: prototype structures

In the chapters of this thesis that cover the theoretical aspects of the project (chapters 3 and 4), as well as in [1] we have seen that power harvesting pn-junctions tend to perform better when wide depletion regions are available; a situation that arises either when the volume around the metallurgical surface is very lightly doped or when junctions are deliberately engineered with an intrinsic region separating the unequivocally p- or n-type regions (pin-junctions).

Within the confines of CMOS technology the following question arises: how do different CMOS-compatible junction types and junction components perform as power scavengers? Notably, in this setting, parameters such as junction depth and doping profile are outside the control of the design engineer. The theory may give some indications based on ideal, computed doping profiles, but in reality the simplest way to find out is to create sample devices in various technologies and physically test them. Use of semiconductor modeling tools such as TCAD can also provide realistic results given reliable information about the fabrication process, but such analysis falls outside the mandate of this project.

6.1.1 Design summary

Given the standardised nature of CMOS pn-junctions and the potentially non-uniform reference beam profile (the light beam used to power the test devices), we can predict that the following parameters will prove to be important when characterising our semiconductor structures for power harvesting capability: Device ID, junction type, functional group, net junction area, net junction perimeter (and therefore also net area/perimeter ratio), footprint and special features. All of these are self-explanatory, but it is noteworthy that we need not mention junction depth, as that is fixed by the technology for each junction type, but we need to mention footprint in order to cope with potentially non-uniform beam profiles. Overall, this allows us to quickly summarise all devices (every individual pn-junction) in a few tables (6.1 and 6.2).

6.1.2 Expected performance

Taking a leaf from corresponding analysis for the modulation case (chapter 5) the functional groups developed for the study of areal and side-wall component contribution to modulation performance can be reused in order to yield results for power harvesting performance. The expectation is that performance will turn out to be a function of both net junction area and net junction perimeter with relatively little interaction between the two components. For more information on the geometry of the test devices, please consult chapter A of the appendix.

Nevertheless, the specifics of the relations linking energy recovery performance to manufacturing technology, junction type and device geometry are likely to be the reverse of what has been seen for modulator performance. Modulator performance has been theoretically shown to improve with abrupt doping profiles and high doping concentrations; a stark contrast to what is needed for high power recovery efficiencies. Thus, scaling down technology is expected to render power recovery on-chip much more challenging unless special processing steps are introduced that will make pn-junctions with large

Table 6.1: Summary of all test device designs containing key information for power scavenger performance evaluation. Legend: Type: junction type. Group: function group, if any. A/P ratio: Area to perimeter ratio. Footprint: nominal design area allocated to the structure, the value shown corresponds to the side of the footprint, which is itself square in shape. Area: Net junction area. Perim: Net junction perimeter.

POWER SCAVENGER DESIGN SUMMARY						
Device ID	**Type**	**Group**	**A/P ratio**	**Footprint**	**Area**	**Perim**
			μm	μm	μm^2	μm
NIN-1	NS	-	1.208	479	75204	62280
NIN-2-pN	pN	-	0.183	479	27878	152064
NIN-2-NS	NS	-	7.317	479	84142	11500
NIN-4	nS	-	16.742	479	206626	12342
NIN-5	NS	-	3.660	479	133848	36566
NIN-6	NS	-	3.660	479	133848	36566
NIN-9	nS	-	0.259	479	78766	304134
SVJ-1	nS	G1	∞	300	11615	0
SVJ-2	nS	G1	∞	300	11615	0
SVJ-3-pN	pN	G2	1.816	300	30628	16865
SVJ-3-NS	NS	-	∞	300	31819	0
SVJ-4-pN	pN	G2	∞	300	6757	0
SVJ-4-NS	NS	-	∞	300	6757	0
SVJ-5	nS	G3	1.562	300	38698	24768
SVJ-6	nS	G3	inf	300	21976	0
SVJ-7	NS	G4	0.316	200	7045	22282
SVJ-8	NS	G4	0.455	200	10615	23316
SVJ-9	NS	G4	0.973	200	16017	16466
SVJ-10	NS	G4	0.455	200	10615	23316
SVJ-11	NS	G4	2.567	200	24524	9554
SVJ-12-3N	3N	G5	0.596	200	13844	23232
SVJ-12-NS	NS	-	5.882	200	24559	4175
TED-1	NS	G1	0.264	495	35123	133004
TED-2	nS	G2	0.285	495	127116	446332
TED-3	nS	G2	9.313	495	199621	21435
TED-4	NS	G1	19.369	495	217978	11254
TED-5-3N	3N	G3	0.342	495	44362	129565
TED-5-NS	NS	-	10.979	495	182688	16640
TED-7-n3	n3	G4	6.827	299	53109	7779
TED-7-3N	3N	-	7.885	299	62497	7926
TED-7-NS	NS	-	8.924	299	72055	8074
TED-8-n3	n3	G4	1.219	299	19587	16068
TED-8-3N	3N	-	1.919	299	46930	24461
TED-8-NS	NS	-	9.356	299	73698	7877
TED-9-3N	3N	G3	1.935	299	47339	24461
TED-9-NS	NS	-	9.579	299	75308	7862

Table 6.2: Special feature table covering all designed devices.

SPECIAL FEATURES	
Device ID	**Special feature**
NIN-6	No passivation
NIN-9	Polysilicon masking
SVJ-1	Perimeter-shading mask
SVJ-2	Perimeter-shading mask
SVJ-3-pN	NW perimeter-shading mask
SVJ-4-pN	NW and p+ perimeter-shading masks
SVJ-6	Perimeter-shading mask
SVJ-8	Rectangular basic cell
SVJ-10	Rectangular basic cell

depletion regions available.

An important note to make here is that the study of some junction types within this chapter serves the purpose of completeness in the sense that their use as power harvesters in the context of CMOS electronics is impractical. The best example of this is the standard well to substrate junction that, when operational, biases the well to a voltage below GND. That is a major reason why diffusion to well junctions are frequently used for power generation in CMOS processes where triple wells are not available (potentially SOI processes could exploit well to substrate junctions but those processes begin to depart from 'old school' standard CMOS territory). Yet these junctions still merit study in their power recovery quality if for no other reason than comparison with 'usable' junction types or perhaps for use in dedicated solar battery chip applications whereby an entire die is sacrificed as a solar battery and feeds other parts of the system through standard bond wires.

6.2 Experimental procedure and results (photo-elements)

This section primarily presents the results extracted from power harvesting mode characterisation tests ran on the modulator devices designed throughout this project. Comments stemming directly from an inspection of the gathered results are provided throughout the section, but a closer look at the interpretation and significance of these will be taken in section 6.4.

The experimental set-up and the protocols used to operate it are also included in this section due to their sheer simplicity. Generally, the experiment used to extract information from our power scavenger devices had simple objectives: obtain an I-V sweep of each device under well-controlled conditions of illumination and alignment. Then, the obtained raw data can be easily processed to yield estimates of such basic metrics of optical power scavengers as the open circuit voltage, the short circuit current, the maximum provided power and the power fill factor.

6.2.1 Experimental set-up, protocols and basic result processing

The experimental set-up was very basic. The emitter-side consisted of nothing more than a powerful, 170 mW optical power-output LED with a central wavelength of 635 nm (Thorlabs - LED635L) connected

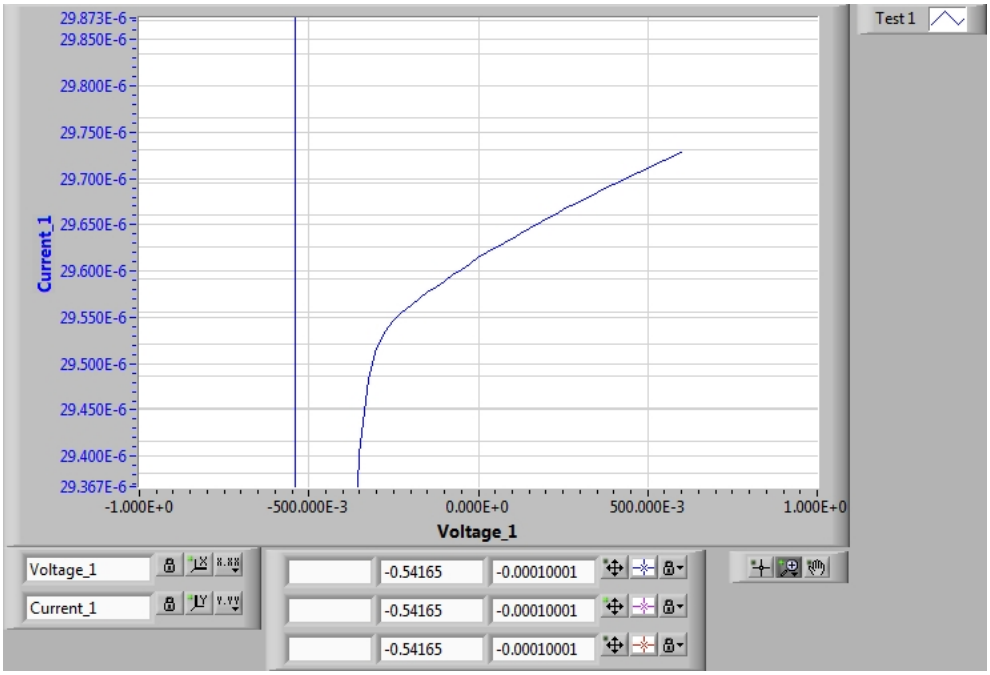

Figure 6.1: Typical 'refined' I-V trace obtained after averaging 20 'raw' voltage sweeps. We have zoomed to the reverse bias region of the I-V curve where we notice the smooth nature of the curve; an indication of low random noise. Axes are in S.I. units.

to an SMU (Keithley - 2602A). The receiver side simply consisted of the photodiode DUTs (devices under test) linked to another SMU (Keithley - 2602A).

Test protocol:

The protocol used to operate the set-up consisted of an alignment and a testing stage. In the alignment stage the target device was connected to the SMU and manually aligned until an acceptable maximum photocurrent value was obtained. Once that was done, a Labtracer (software by Keithley) script was ran and the SMU swept the voltage range between $-0.6\,V$ and $+0.6\,V$ in $25\,mV$ intervals while measuring photocurrent. Each sweep was performed 20 times with $100\,ms$ intervals between spot measurements (step-to-step settling time) and $1\,s$ intervals between sweeps (sweep-to-sweep settling time). The 20 I-V curves obtained were subsequently averaged. This yielded a single, averaged I-V characteristic for each device under illumination and helped eliminate noise. An example of one such averaged I-V sweep can be seen in Figure 6.1.

'Standard illumination':

During both alignment and measurement stages the LED was fed a constant bias current at $350\,mA$. The result was what we shall call 'standard illumination', characterised by an average irradiance of approx. $279.55\,pW/\mu m^2$ (measured) at $635\,nm$ central wavelength with $15\,nm$ wavelength FWHM (full width at half maximum) (both from the datasheet - see Figure 6.2). The LED emission spectrum can be compared to the Silicon absorption spectrum seen in Figure 3.2 on page 65. Evidently Silicon is not very efficient at absorbing that wavelength compared to blue or green light, but as the existence of CMOS imagers amply testifies it is still practical to work at those wavelengths[1]. Table 6.3 summarises

[1]Originally the idea was to test photoresponse at blue and green wavelengths, but time constraints sadly didn't allow

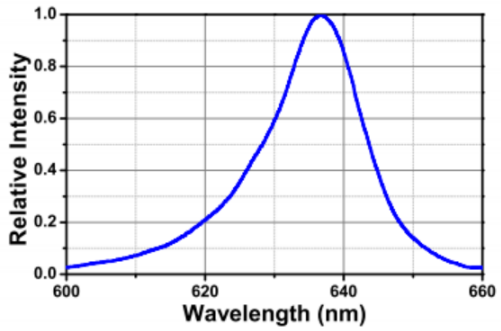

Figure 6.2: Emission spectrum for LED used in our power recovery experiments. Reproduced rom the datasheet as found at http://www.thorlabs.us/ Thorcat/ 22300/ LED635L-SpecSheet.pdf

the illumination conditions used for all the experiments in this chapter.

Table 6.3: Specification of standard illumination. $\overline{E_e}$: average irradiance. $\lambda_{central}$: central wavelength. FWHM: Full width at half maximum.

STANDARD ILLUMINATION STATS		
Parameter	**Value**	**Units**
$\overline{E_e}$	279.55	$pW/\mu m^2$
$\lambda_{central}$	635	nm
FWHM	15	nm

All entries of table 6.3 were determined by running a calibration measurement round using an optical power meter (Thorlabs - PM100D), a calibrated photodiode that responds to the correct wavelengths (Thorlabs - S120C) and a $150\,\mu m$ diameter pinhole (Thorlabs - P150S) that ensures that only a small area of the calibrated photodiode is exposed to the illumination. We begin by setting the wavelength parameter of the power meter at $635\,nm$ which means that the power meter operates under the implicit assumption that all incoming radiation is of $635\,nm$ wavelength. This introduces certain amount of error in the measurement, however repeating the measurement with the assumption of $630\,nm$ wavelength yields very similar values. We therefore can reasonably assume that the errors introduced by the spectral content of the LED illumination will be 'sufficiently small'. Next, we had to align the set-up so that the optical power output of the calibrated photodiode was maximised. This, in combination with the presence of the pinhole, ensures that we measure irradiance under illumination conditions that correspond well to what a properly aligned real device will be exposed to. Finally, irradiance can be easily calculated by dividing the obtained power reading by the area of the pin-hole.

Note: The separation between LED and the calibrated photodiode was kept similar (within approx. $1\,cm$) to the separation between test dies and LED in the region of $10\,cm$.

Note: The calibration run and all measurements in this chapter were taken only after a minimum thermal stabilisation period of 30' duration. Care was taken so that the distance between LED and calibrated photodiode corresponded to the distance between the LED and its device targets on our test chips.

---

that.

Note: In the case of multi-junction cells, each junction was tested with all 'downstream' junctions shorted and all 'upstream' junctions left floating. This was, just like in the experiments concerning modulation performance, a result of the PCB design; a design that allowed for very easy 'downstream' junction shorting but made 'upstream' junction shorting very difficult and potentially unreliable.

As mentioned above, the irradiance under standard illumination takes a value of $279.55\,pW/\mu m^2$, which can be compared to some benchmark values: a) The maximum irradiance allowed to fall on a high optical power meter utilising a Silicon photodiode. b) Typical irradiance values seen by solar cells. Despite the fact that a proper comparison between standard illumination and benchmarks requires taking into account the spectra of each illumination standard the irradiance figures alone are enough to place the differences into perspective.

First we compare our standard illumination irradiance to the maximum irradiance rating of the ThorLabs S121C, Silicon power measurement photodiode[2] of $200\,nW/\mu m^2$ (converted from the $20W/cm^2$ quoted on the datasheet). However, allowances need to be made for the fact that the S121C is covered by a neutral density filter with nominal optical density equal to 2. This means that only about 1 % of the incoming radiation actually reaches the photodetecting surface. Thus, the photodetector is rated to receive up to $2000\,pW/\mu m^2$, or approximately ten times the irradiance to which we subject our CMOS photo-electric elements.

Note: the quoted maximum irradiance value of $2000\,pW/\mu m^2$ is not a damage threshold for the Silicon element. Damage thresholds are functions of exposure time, wavelength (or equivalently incident radiation spectrum), material thickness (more generally geometry), illuminated material area and shape, the degree of transparency of the material, the presence and capabilities of any heat sinks and many more factors and consequently determining them is not a trivial matter [2].

Next, we can compare standard illumination irradiance to values that can be encountered by solar cells on Earth. Using the UK meteorological office's national solar radiation maps[3] we see that the hottest parts of the country can receive $20\,MJ$ of solar energy over an average day in July. Assuming that day and night are of equal duration we compute an average, daily irradiance of approximately $463\,W/m^2$ or equivalently $463\,pW/\mu m^2$. Thus, the irradiance under our standard illumination is below the average irradiance a solar cell on the Isle of Wight can expect to be subjected to on an average day in July.

Key performance metric extraction:

Our key performance metrics were extracted from the raw I-V data by means of a simple MATLAB script. Plots of some obtained I-V curves will be shown as examples, but for the most part data is just tabulated for brevity.

The methods used by the MATLAB script used to extract the relevant information are:

- Open circuit voltage (V_{OC}): Beginning from full reverse bias each pair of adjacent I-V data points[4] is examined. The last pair that consists of members with different signs (a pair we will call the 'last heterosigned pair') is arbitrarily considered to straddle the actual open circuit point. Linear

[2] Whose data-sheet can be found at http://www.thorlabs.de /Thorcat /18300 /S121C-SpecSheet.pdf in full.
[3] Found at http:// www.metoffice.gov.uk /renewables /solar.
[4] I.e. data points that were taken at bias voltage levels differing by the minimum voltage step value of $50\,mV$.

interpolation between the members of the pair is used to compute an estimate for this value. The method has its weaknesses, but should yield 'reasonable' approximations nevertheless.

- Short circuit current (I_{sh}): Directly available by extracting the photocurrent at the known bias voltage equal to zero.
- Maximum power generation (P_{max}): All points are checked for power dissipation. The point with the most negative figure (a.k.a. the higher power generation) is picked and displayed.
- Power fill factor (F_P): Defined and computed as the ratio of $\frac{I_{sh}V_{OC}}{P_{max}}$.

Note: Interestingly, I_{sh} suffers from y-axis uncertainty only whilst V_{OC} suffers from both x- and y-axis uncertainties. Use of suitable pairs of heterosigned I-V data point pairs in combination with linear interpolation was an effort to reduce the uncertainties in V_{OC}, but in the presence of noisy input arriving from some of the weaker power scavengers, (devices that exhibit signals comparable in magnitude to their own noise) much room for error still exists. The presence of data-points corrupted by noise badly enough to cause their sign to change may cause a significant shift in the location of the last heterosigned I-V pair. It thus follows that devices that exhibit weak photocurrents will be disproportionately affected by uncertainty than their stronger counterparts. Nevertheless, both I_{sh} and V_{OC} errors were likely negligible because of the 20-fold averaging performed on the I-V curves obtained from the SMU.

6.2.2 The Ninja family performance

The 'Ninja' family represents a 0.35 micron manufacturing node and therefore is expected to feature the lowest doping concentrations and the smoothest p- to n-type material transitions out of all design families. This is a result of the less fine processing steps involved. Ninja die #1 (NIN1) was the sole die tested for power recovery performance in this design family. Note: All 'upstream junctions' are floating in these experiments (i.e. the p-diffusion is floating when measuring the performance of the N-well on substrate component of device #2 (NIN1-2-NS)).

Ninja die #1 (NIN1)

Results extracted from Ninja die #1 (NIN1) can be seen in table 6.4.

It is noticeable that a photocurrent threshold can be set (e.g. $15\,\mu A$) below which only diffusion-based junctions lie and above which only well-based junctions lie. The junction type-based difference in photocurrent output levels is mirrored and magnified when considering the maximum power output that each device is found capable of providing. This is hinting towards the (expected) possibility that well-based junctions are more efficient power harvesters than diffusion-based junctions. Whether this is indeed the case or not can only be revealed by combining electrical data with geometrical data. This will be carried out in a later section (6.4).

Finally, device Ninja die #1, dev. #2, junction p-diffusion on N-well (NIN1-2-pN) seems to feature an unusually low power fill factor. This is likely to indicate that p-diffusion on N-well junctions (pN) are

Table 6.4: Basic metrics of the devices residing on Ninja die #1 (NIN1) in the power harvesting configuration.

NINJA POWER PERFORMANCE					
DEV	**TYPE**	V_{OC}	I_{sh}	P_{max}	**Fill**
		V	μA	μW	%
NIN1-1	NS	0.4939	29.6140	11.6300	79.5649
NIN1-2-pN	pN	0.4888	3.8003	1.3631	73.3778
NIN1-2-NS	NS	0.4874	23.5680	9.1029	79.2507
NIN1-4	nS	0.5001	12.0420	4.7648	79.1188
NIN1-5	NS	0.4952	29.2030	11.5114	79.5945
NIN1-6	NS	0.4940	27.7290	10.8832	79.4441
NIN1-9	nS	0.4828	5.6335	2.1259	78.1667

'Standard illumination' used: Average irradiance: $279.55\,pW/\mu m^2$. Central wavelength: $635\,nm$. FWHM: $15\,nm$.

inherently less efficient power harvesters than their N-well on substrate (NS) and n-diffusion on substrate (nS) counterparts.

6.2.3 The Svejk family performance

Two dies representing the Svejk family were tested: Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Each die was tested in its entirety within a single measurement run before being removed from the set-up in order to give way to another die. Thus the consistency of experimental conditions between different devices on the same die was maintained to the highest possible level (alignment, temperature, lighting etc.). Note: All 'upstream junctions' are floating in these experiments (as an example the p-diffusion is floating when measuring the performance of the N-well on substrate component of devices #3 and #4 in both Svejk dies (SVJ-3,4-NS)).

Svejk die #1 (SVJ1)

Results extracted from Svejk die #1 (SVJ1) are shown in table 6.5.

In terms of short-circuit current there seems to be a clear dividing line: all N-well on substrate (NS) type junctions or combinations including an N-well on substrate (NS) component show an I_{sh} of above $1\,\mu A$. All other devices show sub-μA values. Similar separation of devices can be seen in terms of maximum power generation (threshold of e.g. $0.3\,\mu W$). Finally, in terms of power fill factor we notice that n-diffusion on substrate (nS) devices tend to underperform both N-well on substrate (NS) and p-diffusion on N-well (pN) counterparts. Interestingly this is in contrast to the findings from our AMS35 die and probably reflects different doping strategies employed in the AMS35 and IBM18 technologies.

Svejk die #2 (SVJ2)

Results extracted from Svejk die #2 (SVJ2) are displayed in table 6.6.

Svejk die #2 (SVJ2) behaves very similarly to Svejk die #1. The high inter-die consistency means

Table 6.5: Basic metrics of the devices residing on Svejk die #1 (SVJ1) in the power harvesting configuration.

SVEJK 1 POWER PERFORMANCE					
DEV	**TYPE**	V_{OC}	I_{sh}	P_{max}	**Fill**
		V	μA	μW	%
SVJ1-1	nS	0.4186	0.1417	0.0424	71.4944
SVJ1-2	nS	0.4309	0.1558	0.0495	73.6949
SVJ1-3-NS	NS	0.4529	2.3920	0.8454	78.0386
SVJ1-3-pN	pN	0.4680	0.2813	0.1042	79.1541
SVJ1-4-NS	NS	0.4394	1.5066	0.5145	77.7177
SVJ1-4-pN	pN	0.4295	0.0639	0.0213	77.5060
SVJ1-5	nS	0.4736	0.5866	0.2120	76.3057
SVJ1-6	nS	0.4532	0.3060	0.1049	75.6150
SVJ1-7	NS	0.4580	1.7049	0.6045	77.4116
SVJ1-8	NS	0.4669	2.5079	0.9175	78.3557
SVJ1-9	NS	0.4706	2.7179	1.0061	78.6627
SVJ1-10	NS	0.4677	2.5263	0.9264	78.4056
SVJ1-11	NS	0.4729	3.0111	1.1223	78.8194
SVJ1-12-NS	NS	0.4727	2.8661	1.0682	78.8499
SVJ1-12-3N	3N	0.4833	0.1283	0.0483	77.8733

'Standard illumination' used: Average irradiance: $279.55\,pW/\mu m^2$. Central wavelength: $635\,nm$. FWHM: $15\,nm$.

Table 6.6: Basic metrics of the devices residing on Svejk die #2 (SVJ2) in the power harvesting configuration.

SVEJK 2 POWER PERFORMANCE					
DEV	**TYPE**	V_{OC}	I_{sh}	P_{max}	**Fill**
		V	μA	μW	%
SVJ1-1	nS	0.4157	0.1441	0.0426	71.1118
SVJ1-2	nS	0.4295	0.1576	0.0499	73.6593
SVJ1-3-NS	NS	0.4516	2.4004	0.8448	77.9359
SVJ1-3-pN	pN	0.4677	0.2901	0.1074	79.1294
SVJ1-4-NS	NS	0.4367	1.4821	0.5016	77.4996
SVJ1-4-pN	pN	0.4285	0.0639	0.0212	77.4371
SVJ1-5	nS	0.4724	0.6094	0.2196	76.2800
SVJ1-6	nS	0.4523	0.3148	0.1075	75.4856
SVJ1-7	NS	0.4562	1.7110	0.6043	77.4084
SVJ1-8	NS	0.4654	2.5834	0.9408	78.2477
SVJ1-9	NS	0.4690	2.7943	1.0303	78.6238
SVJ1-10	NS	0.4651	2.5221	0.9184	78.2916
SVJ1-11	NS	0.4694	2.9638	1.0973	78.8703
SVJ1-12-NS	NS	0.4699	2.8744	1.0655	78.8905
SVJ1-12-3N	3N	0.4812	0.1286	0.0481	77.6870

'Standard illumination' used: Average irradiance: $279.55\,pW/\mu m^2$. Central wavelength: $635\,nm$. FWHM: $15\,nm$.

that all observations that apply to one of the dies also apply to the other.

6.2.4 The Teddy family performance

Two Teddy dies were examined under the power harvesting capability assessment protocol: Teddy die #5 (TED5) and Teddy die #6 (TED6). Much like for Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2), results from each die were collected in a single test-run in order to minimise the effects of relatively easily eliminable sources of error (temperature variation with time, alignment issues introduced by the process of removing and then repositioning the host PCB on the test set-up etc.). Note: All 'upstream junctions' are floating in these experiments (as an example the n-diffusion and triple-well are both floating when measuring the performance of the N-well on substrate component of devices #7 and #8 in both Teddy dies (TED-7,8-NS)).

Teddy die #5 (TED5)

Results extracted from Teddy die #5 (TED5) are shown in table 6.7.

Table 6.7: Basic metrics of the devices residing on Teddy die #5 (TED5) in the power harvesting configuration.

TEDDY 5 POWER PERFORMANCE					
DEV	**TYPE**	V_{OC}	I_{sh}	P_{max}	**Fill**
		V	μA	μW	%
TED5-1	NS	-0.4874	23.1920	8.9996	79.6199
TED5-2	nS	-0.4867	3.1836	1.2212	78.8091
TED5-3	nS	-0.4946	3.1606	1.2551	80.2941
TED5-4	NS	-0.4851	22.9040	8.7926	79.1315
TED5-5-NS	NS	-0.4800	19.9860	7.5926	79.1391
TED5-5-3N	3N	0.4756	-1.0802	0.4055	78.9335
TED5-7-n3	n3	-0.4472	0.3410	0.1196	78.4690
TED5-7-3N	3N	0.4688	-0.6914	0.2574	79.4153
TED5-7-NS	NS	-0.4776	8.2526	3.1103	78.9058
TED5-8-n3	n3	-0.4459	0.1498	0.0523	78.3202
TED5-8-3N	3N	0.4760	-0.6516	0.2459	79.2866
TED5-8-NS	NS	-0.4802	10.4040	3.9371	78.8000
TED5-9-NS	NS	-0.4804	9.1079	3.4648	79.1940
TED5-9-3N	3N	0.4806	-0.7128	0.2724	79.4991

'Standard illumination' used: Average irradiance: $279.55\,pW/\mu m^2$.
Central wavelength: $635\,nm$. FWHM: $15\,nm$.

In terms of short-circuit current there is a clear distinction between NS-type and other junctions. This distinction also applies for maximum generated power values. Suitable thresholds can be easily found that divide the devices accordingly. Notably, all junction types seem to have similar power fill factors in this technology.

Teddy die #6 (TED6)

Results extracted from Teddy die #6 (TED6) are summarised in table 6.8.

Table 6.8: Basic metrics of the devices residing on Teddy die #6 (TED6) in the power harvesting configuration.

TEDDY 6 POWER PERFORMANCE					
DEV	**TYPE**	V_{OC}	I_{sh}	P_{max}	**Fill**
		V	μA	μW	%
TED6-1	NS	-0.4878	24.2820	9.4261	79.5812
TED6-2	nS	-0.4865	3.0550	1.1701	78.7301
TED6-3	nS	-0.4910	2.8273	1.1125	80.2362
TED6-4	NS	-0.4869	24.9340	9.6205	79.2499
TED6-5-NS	NS	-0.4815	21.4920	8.1834	79.0855
TED6-5-3N	3N	0.4773	-1.1665	0.4402	79.0720
TED6-7-n3	n3	-0.4449	0.3139	0.1097	78.5588
TED6-7-3N	3N	0.4674	-0.6554	0.2433	79.4164
TED6-7-NS	NS	-0.4778	8.3866	3.1601	78.8544
TED6-8-n3	n3	-0.4447	0.1422	0.0495	78.2180
TED6-8-3N	3N	0.4772	-0.6921	0.2622	79.3736
TED6-8-NS	NS	-0.4817	11.1860	4.2486	78.8424
TED6-9-NS	NS	-0.4818	9.8029	3.7388	79.1634
TED6-9-3N	3N	0.4834	-0.8067	0.3097	79.4266

'Standard illumination' used: Average irradiance: $279.55\,pW/\mu m^2$. Central wavelength: $635\,nm$. FWHM: $15\,nm$.

The observations that can be made about Teddy die #6 (TED6) also apply to Teddy die # 5 (TED5). Much like the case was for the Svejk type IBM18 dies, our Teddy type UMC13 dies showed very good inter-die consistency.

6.3 Supporting experiments

Within the framework of power-harvesting capability characterisation a few supporting experiments have been performed with the express purpose of determining to what extent the obtained results are reliable. Also they proved helpful in the attempt to understand what factors cause the experiment to behave non-ideally and corrupt the measured values. Below brief descriptions of the aims, test strategies and results of these supporting experiments are shown.

6.3.1 Determination of beam profile

The determination of the beam profile was perhaps the most important supporting experiment carried out. Much like in the case of modulation, the profile of the beam determines whether and what adjustments need to be made when calculating the power harvesting efficiency of devices that feature different footprints. Flat beam profiles indicate that no adjustments are necessary according to size, but all other profiles require calculations to either determine relevant correction factors for devices featuring different

footprint sizes or prove that all such factors are close enough to unity in order to be assumed equal to one.

Set-up and procedure: The set-up used for this experiment was approximately the same as the one used for the main experiment, only ran with a $25 \times 25\,\mu m$ photodiode; the same as was used for the equivalent experiment in the chapter on modulation performance (see B.2.2). The origin (point (0,0)) was set to be a local maximum homed-upon manually. All measured photocurrents were subsequently normalised to the value obtained at the origin. The step size in either direction was determined by the micropositioner attributes at a value of approx. 100 microns. The conventions for what constitutes the x- and y-directions are consistent with the ones used for the corresponding modulation experiment.

Results: Plots of the normalised photocurrent values vs. location are given in Figure 6.3. Interestingly, the plots show that there are at least two local maxima in both x- and y-directions. During quick test runs where the x- and y-directions were swept along longer paths it was observed that the beam profile showed four or more peaks in each direction (results not shown here). The peaks are characterised by very similar peak amplitudes and reasonably flat tops. Specifically, it would seem that the drop in average irradiance up to 200 microns away from any local maximum hardly exceeds 3%. Meanwhile, the difference in maximum irradiance between adjacent peaks seems to lie in the 0-2% range. Interestingly, even within the troughs between successive peaks the collected photocurrent intensity doesn't drop by more than 3-4 percentage units, which is a testimony to a relatively 'flat' irradiance distribution over a large area.

Important note: These profiles will depend on the separation between the photo-transductive element and the LED. However, care was taken to ensure that all dies and the photodiode used for the determination of the beam profile in the present experiment were situated at similar separations from the LED (separation in the region of $\approx 10\,cm$, variation in separation in the sub-cm range). The effectiveness of this method was confirmed during the manual alignment stages performed for each individual device we have tested when we observed the peaks shown in Figure 6.3 at similar intervals (although precise measurements of the inter-peak intervals were not carried out).

These facts allow us to consider that so long as a DUT has been aligned properly to a local maximum peak: a) the size of the device should make a negligible difference to the average amount of irradiance it receives over its entire area and b) any measurement inconsistencies are more likely to occur as a result of DUT alignment on the wrong peak. Furthermore the comparative result errors introduced by homing a DUT on to the wrong local maximum seem to be significantly larger than any such inconsistencies introduced due to the different footprint sizes of various devices.

In conclusion, we shall consider that regardless of misalignment and differing footprint sizes all devices receive the same, constant irradiance throughout their entire areas. Note: in order to render this assumption as close to valid as possible we decided to spend enough time and effort searching for the global irradiance maximum before deeming the alignment point as 'acceptable'. Nevertheless, we can still not full-proof guarantee that we have always managed to successfully home on to the global maximum.

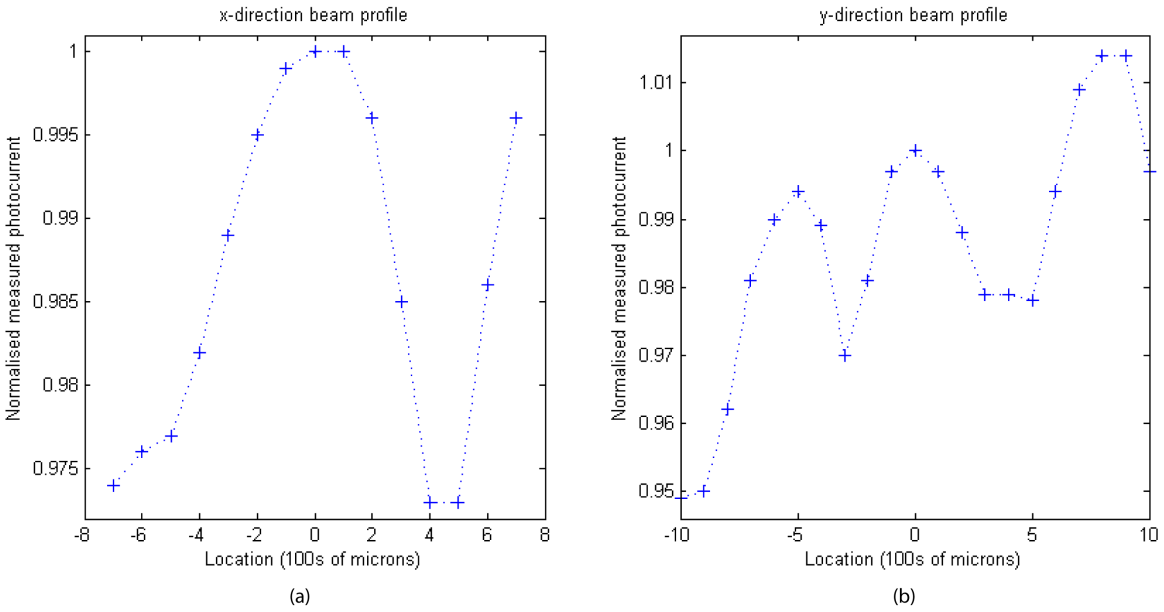

Figure 6.3: Photocurrent intensity as a function of die location vs. light emitter within the context of power recovery experiments. Location is coded in terms of excursions from an arbitrary refernce point in the x- (a) and in the y-directions (b). Photocurrent magnitudes are normalised with respect to the corresponding value at the arbitrary reference location. The origins in each panel represent one and the same point.

6.3.2 LED thermal stabilisation test

A test was ran whereby the photocurrent of a selected photodiode (in this case Teddy die #6, dev. #1 (TED6-1)) was measured at certain time intervals after LED switch-on. The objective was to determine how quickly the LED suffers the effects of thermal quenching and how much they affect performance overall.

Set-up and procedure: The set-up used was in no way different to the main experimental set-up, but the protocol was different. At switch-on a photocurrent measurement was taken. Subsequently, three more measurements were taken at 15', 30' and 2 hr 30' after switch-on.

Results: The results are summarised in table 6.9. The numbers show that after a 30' wait most of the thermal stabilisation process has run its course. For that reason all experiments on power harvesting capability were ran after an initial 30' wait.

Table 6.9: Time course of power-emitting LED thermal stabilisation phenomenon. The entries in the last row show the percentage drop between the spot measurement of their own column vs that of the previous column.

THERMAL STABILISATION TIME COURSE				
Time (in mins)	0	15	30	150
Photocurrent (in μA)	24.10	23.65	23.48	23.38
% drop	-	1.87	0.72	0.43

Additional considerations: Interestingly, a couple of issues arise. First of all, what this experiment

has technically measured is drift over time from switch-on. Despite the fact that the numbers are loosely fitting with the model of an exponential decay, denser measurements would need to be taken in order to confirm that this is indeed the case. If that is indeed confirmed, then the likelihood that what has been measured corresponds to an observation of phenomena dominated by thermal quenching (or indeed a combination of phenomena that arise upon switch-on and decay exponentially with time) will greatly increase. Second, beyond some point, natural ambient temperature changes (e.g. with the time of day) will lead to a long-term drift of the 'asymptotic limit point' towards which the system is tending to converge[5].

6.4 Discussion

Analysis of the power harvesting capability of devices residing on dies of the Ninja, Svejk and Teddy design families will proceed along the following lines: pn-junctions will be considered as consisting of clearly definable areal and side-wall components. On this basis we shall attempt to find how much power is contributed under 'standard illumination' conditions (see table 6.3) by each micron of side-wall and each micron squared of areal junction at their maximum power delivery I-V point. These values we shall call 'junction power coefficients', or simply 'power coefficients'. Studying their values for groups of homotype devices residing on the same physical die and then comparing them between corresponding groups of devices on different physical dies of the same type will be part of this analysis.

The reason for splitting our pn-junctions in such way is to assist practical design in an industrial setting. The different doping properties and geometry of areal and side-wall junction components practically guarantees that they will contribute differently to power recovery. In terms of our project, they will feature different power coefficients p_{fringe} and p_{areal} for side-wall and areal components respectively. This can be qualitatively combined and compared with our theoretical predictions which state that junctions with narrower depletion regions should be weaker optical power scavengers. On the other hand, the layout engineer has access to area and side-wall extent information generally easily and will typically be only interested in finding a way to obtain as much power as possible for given illumination conditions from each square micrometer of chip real estate available. As such, we saw the determination of junction power coefficients as a significant first step towards automating the process of optical power harvester design[6].

Finally, we shall examine the differences in power coefficients between homotype devices residing on dies representing different technological nodes. Any trends in terms of how electronic miniaturisation affects the optical power harvesting capability of each junction type should thus become apparent.

This string of analysis, however, will be preceded by some preparatory work whereby we shall consider inter-die variation in power harvesting capability on a full-device basis, i.e. without taking into consideration the different contributions of areal and side-wall junction components.

Note: The concept of 'fractional error' (otherwise called 'fractional variation') 'F' will be used widely

[5] The point where $\frac{\partial T}{\partial t} = 0$ throughout the entire volume of the LED. This is determined by the amount of heat generation within the LED, the distribution of heat generation, the thermal properties of the materials involved and most crucially, the ambient temperature.

[6] As such this was never designed to be a physics-driven project interested in ab ovo derivations and modelling that competes with professional device-level modelling CAD software or is interested in the details of doping profiles and other minutia of device physics, but rather as an empirically-driven endeavour.

throughout this chapter. This we define as the value yielded by the formula $F = 2\frac{a-b}{a+b}$ i.e. difference over average. The measure is chosen as a more advanced form of $F = a/b$ despite the weak point when it comes to handling value pairs with average equal to zero. This is done in order to achieve a magnitude-commutative relation as exchanging a and b will merely change the sign of F and not its magnitude.

6.4.1 Inter-die variation

We can examine inter-die variation only for the Svejk (SVJ) and the Teddy (TED) die pairs since only a single representative of the Ninja (NIN) family was successfully tested. Out of all the key metrics, in this section we will concentrate on maximum power delivery capability and compare discrepancies in that metric between homologous designs residing on different dies. We shall begin by showing lower level results (comparing individual device pairs), then refine it (compare different junction types as sets) and then reach the final overview (Show a single table summarising key information).

Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2)

If we compare the maximum power delivery capability pair-wise for all devices residing on Svejk die #1 (SVJ1) and Svejk die # (SVJ2), we obtain table 6.10. If we then group devices by junction type and average their fractional variations we obtain table 6.11.

Table 6.10: Inter-die comparison of devices residing on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2), operating in power harvesting mode. The first column contains just the device identifier. P_{max} stands for maximum power delivery and F stands for fractional difference between corresponding devices according to the formula: $F = 2\frac{a-b}{a+b}$. The horizontal line splitting the data in the table in two separates devices with different footprint sizes (large devices above). Standard illumination was used (see table 6.3).

INTER-DIE COMPARISON: SVEJKS				
		SVJ1	**SVJ2**	
DEV	**TYPE**	P_{max}	P_{max}	**F**
		μW	μW	%
1	nS	0.0424	0.0426	-0.47
2	nS	0.0495	0.0499	-0.80
3	NS	0.8454	0.8448	0.07
3	pN	0.1042	0.1074	-3.02
4	NS	0.5145	0.5016	2.54
4	pN	0.0213	0.0212	0.47
5	nS	0.2120	0.2196	-3.52
6	nS	0.1049	0.1075	-2.45
7	NS	0.6045	0.6043	0.03
8	NS	0.9175	0.9408	-2.51
9	NS	1.0061	1.0303	-2.38
10	NS	0.9264	0.9184	0.87
11	NS	1.1223	1.0973	2.25
12	NS	1.0682	1.0655	0.25
12	3N	0.0483	0.0481	0.41

We notice how different junction types show very different degrees of variation. Notably, n-diffusion

Table 6.11: Average fractional variation ($F = 2\frac{a-b}{a+b}$) between maximum power delivery capabilities of devices residing on different dies ($\overline{F}$). The number of devices considered for each type of junction are shown in the last column. Standard illumination was used (see table 6.3).

SVJ JUNCTION TYPES		
TYPE	$\overline{F}$	**n**
	%	
nS	-1.81	4
pN	-1.28	2
3N	0.41	1
NS	0.14	8

on substrate (nS) and p-diffusion on N-well (pN) junctions vary by more than their triple well on N-well (3N) and N-well on substrate (NS) counterparts. Thus, diffusion-based junctions vary by more than well-based junctions.

Some of the variation can be attributed to measurement noise and some due to differences that lie within the Silicon, but overall we observe rather low variation.

Teddy die #5 (TED5) and Teddy die #6 (TED6)

A pair-wise comparison of each junction on Teddy die #5 (TED5) and Teddy die #6 (TED6) yields table 6.12 whilst grouping devices by junction type leads to the results seen in table 6.13.

Table 6.12: Inter-die variation between corresponding devices on Teddy die #5 (TED5) and Teddy die #6 (TED6). The first column shows the device identifier. P_{max} stands for maximum power. F represents the fractional difference between devices under the formula: $F = 2\frac{a-b}{a+b}$. Standard illumination was used (see table 6.3).

INTER-DIE COMPARISON: TEDDIES				
		TED5	**TED6**	
DEV	**TYPE**	P_{max}	P_{max}	**F**
		μW	μW	%
1	NS	8.9996	9.4261	-4.63
2	nS	1.2212	1.1701	4.27
3	nS	1.2551	1.1125	12.05
4	NS	8.7926	9.6205	-8.99
5	NS	7.5926	8.1834	-7.49
5	3N	0.4055	0.4402	-8.21
7	NS	3.1103	3.1601	-1.59
7	3N	0.2574	0.2433	5.63
7	n3	0.1196	0.1097	8.63
8	NS	3.9371	4.2486	-7.61
8	3N	0.2459	0.2622	-6.42
8	n3	0.0523	0.0495	5.50
9	NS	3.4648	3.7388	-7.61
9	3N	0.2724	0.3097	-12.82

The UMC13 die family shows itself to be overall less well-behaved than the IBM18 family. Whereas

Table 6.13: Average inter-die variation of devices representing the same junction types ($\overline{F}$, where $F = 2\frac{a-b}{a+b}$.). The last column shows how many devices have been taken into consideration when generating the average data. Standard illumination was used (see table 6.3).

TED JUNCTION TYPES		
TYPE	$\overline{F}$	**n**
	%	
nS	8.16	2
n3	7.07	2
3N	-5.45	4
NS	-6.32	6

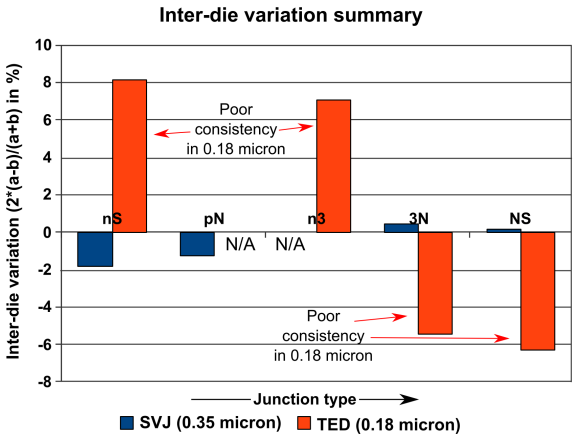

Figure 6.4: Inter-die variation across technologies, according to junction type. The higher the absolute value of each bar, the worse the inter-die consistency. Device types showing particularly good or bad inter-die consistencies have been explicitly labeled. Unavailable data is explicitly shown as 'N/A'. Device type legend: nS: n-diffusion on substrate. pN: p-diffusion on N-well. n3: n-diffusion on triple well. 3N: triple well on N-well. NS: N-well on substrate.

our IBM18, 'Svejk' dies showed inter-die variation of less than 2 % for all junction types, our UMC13, 'Teddy' dies showed variations roughly in the $5 - 8$ % range. Given the generally larger size of devices on the UMC13 die and the finer pitch of the UMC13 technology this is a surprising result which we may attribute to different manufacturing techniques utilised in the two technologies. Overall, inter-die consistency in the UMC13 die was rather disappointing. Note: there seems to be only a small differnce between well- and diffusion-based junctions in the UMC13 technology.

Conclusions

Overall, we found out that the IBM18 technology shows better inter-die consistency than the UMC13 node, and that only in the IBM18 technology well-based junctions behave better than diffusion-based ones. Unfortunately, these conclusions are based on small numbers of dies. Ideally, data on fractional difference distribution could be obtained from many exemplaries and statistical significance tests then carried out. All key data related to inter-die variation is summarised in Figure 6.4.

6.4.2 Effects of geometric structure on power harvesting capability

In this subsection we shall attempt to extract power coefficients for each group of junctions belonging to the same type of device and residing on the same die. We shall begin our analysis from the 0.35 micron Ninja die and continue towards smaller feature sizes with the 0.18 micron Svejk (SVJ) family and finally the 0.13 micron UMC13 family. It must be stressed that homotype junctions residing on heterotype devices will be considered as different types of junctions for the purposes of extracting power coefficients. This is in order to avoid pooling of data from junctions that may potentially experience significant performance differences due to the presence of different dopant environments in their vicinity. The matter of whether such effects can be safely ignored or not will be discussed in the subsection dedicated to understanding the differences between different junction types (6.4.3).

The extraction of power coefficients will be done by either pairing homotype junctions residing within homotype devices (when device groups only consist of two devices) or linear fits (for larger device groups). These pairings or linear fits will be carried out on the basis of a model that follows the function:

$$P_{max}(A, P) = p_{areal}A + p_{fringe}P \tag{6.1}$$

where $P_{max}(A, P)$ which will represent maximum delivered power, A the net[7] extent of the areal junction component, P the net length of the side-wall junction component as illustrated in Figure 6.5, and p_{areal}, p_{fringe} will be the coefficients that are being sought. The units for p_{areal} and p_{fringe} are given as W/m^2 and W/m accordingly in terms of S.I. although we shall use the more practical $pW/\mu m^2$ and $pW/\mu m$ for clarity.

Similar results could be extracted for short circuit current, for example, but not for open circuit voltage as open circuit voltage contributions from areal and side-wall components of a junction cannot be simply linearly added. We concentrate on delivered power only, however, as it is considered to be by far the most relevant design parameter.

Ninja die #1 (NIN1):

The Ninja design family features seven usable devices that are grouped as follows:

- Pure N-well on substrate (NS): 3 items.
- n-diffusion on substrate (nS): 2 items.
- p-diffusiono n N-well (pN): 1 item.
- N-well on substrate (NS) in p-diffusion on N-well on substrate (pNS): 1 item.

From the available devices we can form two pairings. This is because two of the N-well on substrate (NS) junctions are geometrically identical and only differ from one another by means of the presence or absence of the passivation layer atop them. We can estimate the power coefficients for Ninja die #1,

[7] 'Net' meaning corrected for areas shaded by metallisation overhanging the device.

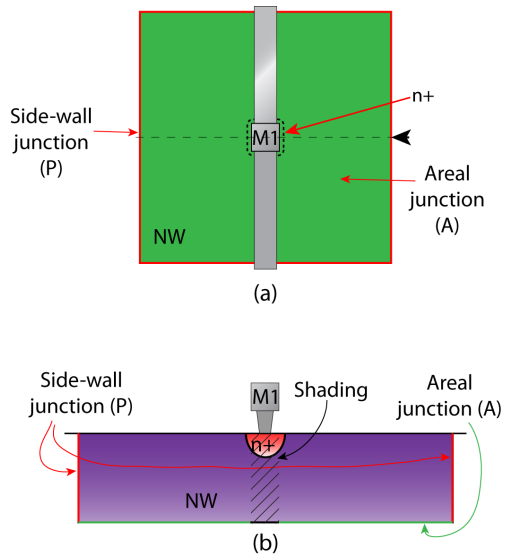

Figure 6.5: Illustration of the concepts of net areal (represented by the green area in (a) and green lines in (b)) and side-wall (shown as red lines) junction components. (a) Top view of an illustrative, idealised pn-junction with a single metal contact to its N-well terminal and the supply line to that contact. (b) Cross-section of the illustrative pn-junction in (a) along the line implied by the black arrowhead in (a). Noteworthy features: (i) The idealised pn-junction sample has very clearly defined areal and side-wall components. (ii) The side-wall component is interrupted by the presence of overhanging metal as seen in (a). (iii) The areal component is interrupted by the presence of overhanging metallisation, but not overhanging diffusion regions (or in general other doped regions) as seen in (b). Abbreviations: NW: N-well. M1: Metal layer 1. n+: n-diffusion.

dev. #6 (NIN1-6) (the device lacking passivation) by extrapolating from the coefficients extracted from Ninja die #1, dev. #5 (NIN1-5). This is implemented by multiplying the coefficients of Ninja die #1, dev. #5 (NIN1-5) by the ratio of the short-circuit photocurrents of Ninja die #1, dev. #6 (NIN1-6) over Ninja die #1, dev. #5 (NIN1-5) (a factor of approx. 0.950). It is rather counter-intuitive that a device missing a physical layer should exhibit lower photocurrent. This hints towards the possibility that either the underlying Silicon in each device has different average properties over the entire extent of each device or could also be the result of an error during the alignment stage of the experiment, or simply be a manifestation of noise.

The results of various device pairings are shown in table 6.14. Apart from the power coefficients corresponding to each formal device grouping, we have added a couple of pairings that link homotype junctions residing in heterotype devices. The purpose is to examine the differences in extracted power coefficients. Note: the p-diffusion on N-well (pN) junction in Ninja design, device #2 (NIN-2) was left floating during measurements on the N-well on substrate (NS) part of the junction.

Extracted power coefficient data seems to suggest that indeed well-based pn-junctions are far more efficient power harvesters than diffusion-based junctions with both areal and fringe coefficients clearly higher than in the well-based devices. Strikingly, the fringe coefficient of the diffusion-based junction pair Ninja die #1, devs. #4 and #9 (NIN1-4,9) is exceptionally small. This could be an indication that perhaps the areal component of diffusion-based junctions completely overshadows the fringe component (as shown in chapter 3 diffusion junctions are expected to have much thinner and less significant side-walls than well-based counterparts). In other words, a square micron of n-diffusion on substrate (nS)

Table 6.14: Maximum power output normalised by junction area or perimeter ('power coefficients') for different junction pairings from the Ninja die #1 (NIN1) die. DEV. PAIR: identifiers of the paired devices used to extract the coefficients. Host type: the full structure of the device hosting each junction. P_{areal}: areal power coefficient. P_{fringe}: side-wall power coefficient. The horizontal line splitting the table data in two groups separates results arising from groupings involving devices residing on homotype hosts from those arising from groupings of devices on heterotype hosts. Standard illumination was used (see table 6.3).

POWER COEFFICIENT SUMMARY					
DIE	**DEV. PAIR**	**Type**	**Host type**	P_{areal}	P_{fringe}
				$pW/\mu m^2$	$pW/\mu m$
NIN1	1,5	NS	NS	52.21	123.69
	6	NS	NS	49.58	117.45
	4,9	nS	nS	23.00	1.03
	1,2-NS	NS	NS/pNS	123.88	37.15
	2-NS,5	NS	NS/pNS	73.32	46.42

junction will behave very similarly regardless of whether it is located at the centre of a large diffusion region or whether it includes a large border with a different semiconductor region. More measurements would be required to confirm or reject this hypothesis.

We also note that power coefficients taken from the pairings between N-well on substrate (NS) junctions forming their own device and N-well on substrate (NS) junctions within a p-diffusion on N-well on substrate (pNS) device are very different from those taken from the pure N-well on substrate (NS) device pairing[8]. This hints towards the possibility that the electro-optical properties of N-well on substrate (NS) devices are heavily affected by the presence of other dopants in the region.

Finally, as expected, all coefficients are positive, which means that both areal and fringe components of each junction contribute a certain amount of power to the whole junction complex.

Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2):

The Svejk (SVJ) design family consists of 15 usable pn-junctions. These are grouped as follows:

- n-diffusion on substrate (nS): 4 items.
- p-diffusion on N-well (pN): 2 items.
- Triple well on N-well (3N) in triple well on N-well on substrate (3NS): 1 item.
- Pure N-well on substrate (NS): 5 items.
- N-well on substrate (NS) in p-diffusion on N-well on substrate (pNS): 2 items.
- N-well on substrate (NS) in triple well on N-well on substrate (3NS): 1 item.

[8]In the case of the pNS type device I-V sweep measurements were taken with the p-diffusion region left floating

Possible photoelectric element configurations also include multi-part pn-junctions such as p-diffusion on N-well on substrate (pNS) (2 items) and tripple well on N-well on substrate (3NS) (1 item). These configurations have been studied in this work, but were found to exhibit very erratic behaviour that defied all our analysis efforts (from experimental results not included in this thesis). As such, results from these structures are omitted from this study.

We will carry out device pairings or linear fits in order to extract power coefficients on n-diffusion on substrate (nS), p-diffusion on N-well (pN), pure N-well on substrate (NS) and N-well on substrate (NS) in p-diffusion on N-well on substrate (pNS) devices. Obtained power coefficients are summarised in table 6.15. Notes: a) Three out of the four n-diffusion on substrate (nS) type junctions on the Svejk (SVJ) design offer no exposed perimeter. Therefore the results pertaining to the side-wall power coefficient may be less reliable than their areal coefficient counterpart. b) Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) offer the same exposed areal and perimeter junction components, but are discernible by the difference they show in the extent of their n-diffusion regions underneath the metal mask (for an image, see 5.12 on page 200). c) N-well on substrate (NS) in p-diffusion on N-well on substrate (pNS) junctions only offer an exposed areal junction component to the incoming light whilst all their perimeter remains under shadow. Thus, only an areal coefficient can be defined for them. This is not included in the table below and is omitted as a result. d) Out of the available two p-diffusion on N-well (pN) devices one of them only offers an exposed areal component with its entire perimeter under shadow. As such the value of the side-wall power coefficient may be less reliable than its areal counterpart. e) All 'upstream' junctions were left floating during measurements.

Table 6.15: Maximum power output normalised by junction area or perimeter ('power coefficients') for the dies of the Svejk (SVJ) family. The second column displays the identifier numbers of all devices that were used in the grouping to which power coefficient values in the last two columns correspond. Standard illumination was used (see table 6.3).

POWER COEFFICIENT SUMMARY					
DIE	**DEV. PAIR**	**Type**	**Host type**	P_{areal}	P_{fringe}
				$pW/\mu m^2$	$pW/\mu m$
SVJ1	7,8,9,10,11	NS	NS	39.66	19.60
	1,2,5,6	nS	nS	4.48	1.56
	3,4	pN	pN	3.15	0.45
SVJ2	7,8,9,10,11	NS	NS	38.85	20.34
	1,2,5,6	nS	nS	4.57	1.73
	3,4	pN	pN	3.14	0.67

The root-mean square error errors (RMSE) associated with the fits performed on the Svejk (SVJ) devices are summarised in table 6.16. The errors are given in nW. Note: power output levels for N-well on substrate (NS) devices on the Svejk (SVJ) platform is typically in the hundreds of nW (500 to 1000 nW) whilst nS devices generate maximum power in the region of tens of nW (approx. 30 to 100 nW).

The inter-die fractional and absolute variation between power coefficients is summarised in table 6.17.

Pairings between various pure N-well on substrate (NS) devices can also be carried out in order to check for the consistency of the group fit and make any gross deviations obvious. For both Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) relevant data is shown in table 6.18.

Table 6.16: Fit RMSD errors corresponding to the N-well on substrate (NS) and n-diffusion on substrate (nS) type junctions of the Svejk (SVJ) dies.

POWER COEFFICIENT FIT ERRORS		
DIE	**DEV. PAIR**	**Error**
		nW
SVJ1	7,8,9,10,11	81.85
	1,2,5,6	8.39
SVJ2	7,8,9,10,11	94.52
	1,2,5,6	9.22

Table 6.17: Fractional (F) and absolute (Δ) inter-die variation of maximum power output normalised by junction area or perimeter ('power coefficients') for Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Standard illumination was used (see table 6.3).

POWER COEFFICIENT VARIATION							
DIE	**DEV. PAIR**	**Type**	**Host type**	F_{areal}	F_{fringe}	Δ_{areal}	Δ_{fringe}
				%	%	$pW/\mu m^2$	$pW/\mu m$
	7,8,9,10,11	NS	NS	2.06	-3.71	0.81	0.74
SVJ1,2	1,2,5,6	nS	nS	-1.88	-10.56	-0.09	-0.17
	3,4	pN	pN	0.47	-38.54	0.01	-0.22

The fractional difference between power coefficient pairs is given by the formula $2 \cdot \frac{a-b}{a+b}$, a represents Svejk die #1 (SVJ1) coefficient and b represents Svejk die #2 (SVJ2) coefficient.

Table 6.18: Maximum power output normalised by junction area or perimeter ('power coefficients') values obtained by pairing pure N-well on substrate (NS) type junctions residing on Svejk (SVJ) dies (Svejk design, dev. #7 (SVJ-7) to Svejk design, dev. #11 (SVJ-11)) in a few, sample combinations. Standard illumination was used (see table 6.3).

SVJ POWER COEFFICIENT SUMMARY				
DIE	**DEV. PAIR**	**Type**	P_{areal}	P_{fringe}
			$pW/\mu m^2$	$pW/\mu m$
SVJ1	7,8	NS	87.86	-0.65
SVJ1	7,10	NS	90.61	-1.52
SVJ1	8,9	NS	42.03	20.21
SVJ1	9,10	NS	41.30	20.93
SVJ1	9,11	NS	35.36	26.71
SVJ2	7,8	NS	95.11	-2.95
SVJ2	7,10	NS	88.21	-0.77
SVJ2	8,9	NS	42.94	20.80
SVJ2	9,10	NS	44.80	18.99
SVJ2	9,11	NS	32.80	30.67

Given the data obtained so far we reach a number of conclusions, first of which is that result consistency between dies is generally good in so much as no corresponding power coefficients differ by more than a single unit from one another. This would indicate that the consistency both of the Silicon and the test set-up is most likely to be of high quality. However, the weak diffusion-based junctions show the most considerable fringe power coefficient fractional discrepancies. It is possible to explain[9] such behaviour by assuming that noise sources that do not scale with the intensity of received light dominate in the set-up and combining this with the facts that diffusion-based junctions are weak respondents to illumination and feature shallow and generally meagre side-walls. No direct noise level measurements were taken under different illumination conditions (in retrospect this would have been a useful endeavour).

All calculated power coefficients were positive as table 6.15 attests, but when we decided to further investigate and computed power coefficients for various pairings of N-well on substrate (NS) devices on our Svejk family dies (SVJ), some negative coefficients appeared. The reasons for this behaviour are unclear, but we can speculate that Svejk design, device #7 (SVJ-7) specimens may exhibit unusual behaviour compared to their larger counterparts because they are designed with basic cells of exactly 1×1 minimum design width areas. This may mean that the doping profiles in such tiny basic cells may be affected by some systematic error introduced by the fact that the technology is being pushed to its limits. In effect, we might be observing a 'blurring' of the boundaries between what we consider to be areal and what we consider to be side-wall junction. This explanation, however, is purely speculative. Further study is required in order to understand exactly what causes this phenomenon and warns against using devices with very small basic cells for the purposes of understanding the contribution of various CMOS junction components to power generation. Note: this areal-side-wall boundary 'blurring' effect, if indeed the cause of this behaviour, is expected to affect any devices with a large percentage of areal junction within one design length of a boundary.

Another noteworthy event concerns the pairings between devices Svejk design, dev. #7 (SVJ-7) through Svejk design, dev. #11 (SVJ-11). There seem to be consistent differences between pairings carried out between devices with different basic cell sizes. If we rank Svejk (SVJ) N-well on substrate (NS) type devices by basic cell size the obtained ranking order is: $7 < 8 = 10 < 9 < 11$ where the numbers represent device identifier numbers. Thus, pairings between Svejk design, devs. #7 and #8 (SVJ-7,8) and Svejk design, devs. #7 and #10 (SVJ-7,10) form the set of pairings between devices with small and ultra-small basic cells (the USS set). Similarly pairings between Svejk design, devs. #8 and #9 (SVJ-8,9) and Svejk design, devs. #9 and #10 (SVJ-9,10) form the set of pairings between devices with small and medium basic cells (The SM set). Finally, the Svejk design, devs. #9 and #11 (SVJ-9,11) pairing alone constitutes the pairing between devices with medium and large basic cells (the ML set). If we consider the power coefficients for each of these sets separately, we see that there seem to be significant differences between both areal and side-wall power coefficients between sets with comparatively very small discrepancies between pairings within the same set. This phenomenon is an extension to the issue of negative side-wall power coefficients and we attempt to address it in section 6.5. The plot resulting from pinpointing the areal and fringe power coefficients on a Cartesian plane is shown in 6.6.

A clear trend develops whereby we see that the larger the basic cell, the more 'positive' the fringe power coefficient and the smaller the areal power coefficient, in essence turning table 6.18 into an image. We note how even devices with larger basic cells seem to show device size-dependent power coefficients.

[9] Not necessarily the only possible explanation, but one that seems plausible nonetheless.

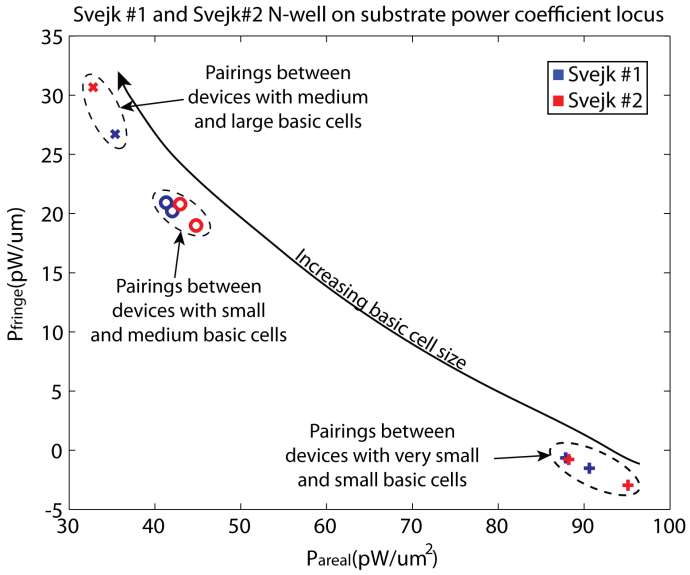

Figure 6.6: Extracted maximum power output normalised by junction area or perimeter ('power coefficients') for N-well on substrate device pairings where the devices involved feature similar basic cell sizes. Each colour represents a die. Blue: Svejk die #1 (SVJ1), Red: Svejk die #2 (SVJ2). Each symbol represents a certain pairing category: +: Very small to small (Svejk deisgn, devs. #7, #8 and #10 (SVJ-7,8,10)). o: Small to medium (Svejk design, devs. #8, #9 and #10 (SVJ-8,9,10)). x: Medium to large (Svejk design, devs. #9 and #11 (SVJ-9,11)). The black arrow roughly indicates the presumed trend that these data points show; i.e. that as basic cell size increases extracted power coefficient values tend towards their 'real' values. See text for more detail. Standard illumination was used (see table 6.3).

Teddy die #5 (TED5) and Teddy die #6 (TED6):

The Teddy (TED) design family consists of 14 usable junctions that can be grouped as follows:

- n-diffusion on substrate (nS): 2 items.
- n-diffusion on triple well (n3): 2 items.
- Pure N-well on substrate (NS): 2 items.
- triple well on N-well (3N) in p-diffusion on triple well on N-well on substrate (p3NS): 2 items.
- N-well on substrate (NS) in p-diffusion on triple well on N-well on substrate (p3NS): 2 items.
- N-well on substrate (NS) in triple well on N-well on substrate (3NS): 2 items.
- Triple well on N-well (3N) in triple well on N-well on substrate (3NS): 2 items.

By pairing devices we can obtain estimates of power coefficients for a multitude of junction types, discriminated by not only by their type, but also by their environment i.e. the presence or absence of other dopants in their immediate vicinity. We obtain table 6.19. Note: All 'upstream' junctions were left floating during measurements.

Note: The N-well on substrate components of devices #5 and #9 (TED-5-NS and TED-9-NS), as well as the n-diffusion on triple-well junctions of devices #7 and #8 (TED-7-n3 and TED-8-n3) are omitted as they yielded results very widely outside the expected and were classed as clear outliers.

Table 6.19: Maximum power output normalised by junction area or perimeter ('power coefficients' - P_{areal} and P_{fringe}) for Teddy die #5 (TED5) and Teddy die #6 (TED6) summary. The table also indicates junction type (col. 2) and host device type (col. 3). Standard illumination was used (see table 6.3). The N-well on substrate components of devices #5 and #9 (TED-5-NS and TED-9-NS), as well as the n-diffusion on triple-well junctions of devices #7 and #8 (TED-7-n3 and TED-8-n3) are omitted as they yielded results very widely outside the expected and were classed as clear outliers.

POWER COEFFICIENT SUMMARY					
DIE	**DEV. PAIR**	**Type**	**Host type**	P_{areal}	P_{fringe}
				$pW/\mu m^2$	$pW/\mu m$
TED5	1,4	NS	NS	37.35	57.8
	7,8	NS	n3NS	61.39	11.85
	5,9	3N	3NS	6.73	-1.88
	7,8	3N	n3NS	-0.13	16.17
	2,3	nS	nS	6.18	0.98
TED6	1,4	NS	NS	41.04	60.03
	7,8	NS	n3NS	61.16	11.85
	5,9	3N	3NS	7.21	-1.29
	7,8	3N	n3NS	0.06	15.07
	2,3	nS	nS	5.46	1.07

Inter-die discrepancies between power coefficients can also be calculated; both absolute and fractional. These are summarised in table 6.20.

Table 6.20: Fractional (F) and absolute (Δ) inter-die variation of maximum power output normalised by junction area or perimeter ('power coefficients') for Teddy die #5 (TED5) and Teddy die #6 (TED6). Standard illumination was used (see table 6.3).

POWER COEFFICIENT VARIATION							
DIE	**DEV. PAIR**	**Type**	**Host type**	F_{areal}	F_{fringe}	Δ_{areal}	Δ_{fringe}
				%	%	$pW/\mu m^2$	$pW/\mu m$
TED5,6	1,4	NS	NS	-9.4	-3.79	-3.68	-2.23
	7,8	NS	n3NS	0.38	-23.42	0.23	-2.48
	5,9	3N	3NS	-6.88	37.67	-0.48	-0.6
	7,8	3N	n3NS	592.88	7.1	-0.19	1.11
	2,3	nS	nS	12.44	-8.99	0.72	-0.09

The fractional difference between power coefficient pairs is given by the formula $2 \cdot \frac{a-b}{a+b}$, a represents Teddy die #5 (TED5) coefficient and b represents Teddy die #6 (TED6) coefficient.

These results lead to certain conclusions and observations, beginning with the fact that the consistency between homologous power coefficients on different dies is good either in terms of absolute values (particularly for coefficients with smaller magnitudes) or in terms of fractional differences (coefficients with larger magnitudes). This could be explained if noise in the determination of the power coefficients arises from the combination of a number of sources that do not scale with the magnitude of the coefficient (keeps fractional differences high between coefficients of low magnitude) and a number of sources that do

(keeps absolute differences between coefficients of high magnitude high), but understanding specifically what processes cause such behaviour should be the target of further study.

Strikingly, we obtain with consistency between dies some negative coefficients. The devices involved, namely Teddy design, devs. #5 and #9 (TED-5,9) are multi-junction. What might be causing significant distortions in the behaviour of what is technically a generic N-well hosting a triple-well (or triple-well plus n-diffusion) may be the different distributions of these extra doped regions on top of their host N-well basin. If indeed the distribution of additional doped regions plays a significant role in determining the behaviour of the host N-well, then even multi-junction devices with identical N-wells in both shape and size and extra doped regions inside the well with the same areal and perimeter junction component extents may still cause the underlying N-well to behave differently between the devices. This explanation might potentially also apply to the outliers mentioned previously. The present study is simply insufficient in order to determine the exact behaviour of N-well on substrate (NS) junctions underlying triple well on N-well (3N) and n-diffusion on triple well on N-well (n3N) junctions.

The negative areal coefficient obtained from the triple-well on N-well junctions of devices #7 and #8 on die #5 (TED5-7,8-3N) may be explained as a simple case of noise tipping a very small 'real' value into a negative 'real + noise' value.

Summary and conclusions

All extracted power coefficients are summarised for convenience in table 6.21. Figure 6.7 summarises the power coefficients for the most 'interesting' devices.

Figure 6.7 shows how the geometry of a device can affect power recovery performance. For each individual die and each junction type we can see the relation between the power recovery capabilities of a square micron of areal junction vs. the contribution from each micron of side-wall. We immediately observe that in some devices the areal component dominates power output whist in other cases it is the side-wall that makes the major contribution. How this relates to each junction type and each manufacturing technology will be discussed farther down this section. Nevertheless we can clearly see how power harvester layouts need to be adjusted so that the stress falls on either large contiguous areal junction components or long side-walls (depending on junction type and manufacturing technology) in order to obtain the most out of a harvester device of constrained area.

Summarising the section on the influence of geometry on power recovery performance we have reached a number of conclusions characterised by varying degrees of confidence:

- Basic cells where large areas are in close proximity to a junction border may lead to the blurring of boundaries between areal and side-wall junction components. Small basic cells, such as the ones used throughout our designs, are particularly vulnerable to this effect. This phenomenon probably pervades the entire power coefficient result table, but was made most clear in the various pairings between the Svejk design, devs. #7, #8, #9, #10 and #11 (SVJ-7,8,9,10,11) devices. Ideally, any pairing between devices belonging to the aforementioned set should have provided the same result, but when grouping pairings by the size of the basic cells of the paired devices, differences immediately appear.
- Inter-die consistency can be considered reasonably good either in terms of fractional or in terms

Table 6.21: Summary of all extracted maximum power output normalised by junction area or perimeter values ('power coefficients' - P_{areal} and P_{fringe}). Shown are the die and identifier pairs (second column) of the grouped junctions along with their type and the full structure of their host device (fourth column). Areal and fringe power coefficients are marked with the letter 'P' and the appropriate subscript. Standard illumination was used (see table 6.3).

POWER COEFFICIENT SUMMARY					
DIE	**DEV. PAIR**	**Type**	**Host type**	P_{areal}	P_{fringe}
				$pW/\mu m^2$	$pW/\mu m$
NIN1	1,5	NS	NS	52.21	123.69
	6	NS	NS	49.58	117.45
	4,9	nS	nS	23.00	1.03
SVJ1	7,8,9,10,11	NS	NS	39.66	19.60
	1,2,5,6	nS	nS	4.48	1.56
	3,4	pN	pN	3.15	0.45
SVJ2	7,8,9,10,11	NS	NS	38.85	20.34
	1,2,5,6	nS	nS	4.57	1.73
	3,4	pN	pN	3.14	0.67
TED5	1,4	NS	NS	37.35	57.80
	7,8	NS	n3NS	61.39	11.85
	5,9	3N	3NS	6.73	-1.88
	7,8	3N	n3NS	-0.13	16.17
	2,3	nS	nS	6.18	0.98
TED6	1,4	NS	NS	41.04	60.03
	7,8	NS	n3NS	61.16	11.85
	5,9	3N	3NS	7.21	-1.29
	7,8	3N	n3NS	0.06	15.07
	2,3	nS	nS	5.46	1.07

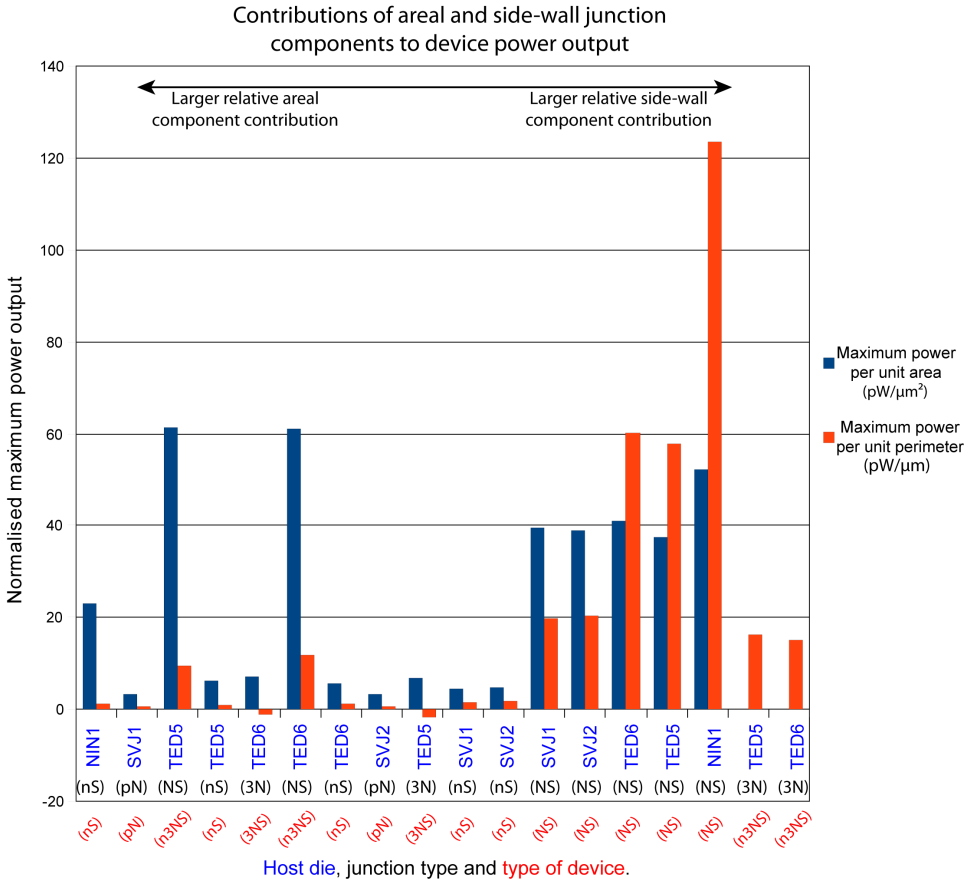

Figure 6.7: Normalised maximum power recovery per unit area (P_{areal}) and per unit perimeter (P_{fringe}) for each junction type on each die. Results are sorted by the ratio of $|P_{areal}|/|P_{fringe}|$ so that the device types that are more 'area-dominated' appear towards the left and those that are more 'fringe-dominated' appear towards the right. Device type legend: nS: n-diffusion on substrate. pN: p-diffusion on N-well. NS: N-well on substrate: n3: n-diffusion on triple well. 3N: triple well on N-well.

of absolute error depending on the magnitude of the power coefficient under study. When power coefficients are small in magnitude they seem strongly affected by uncertainty factors that do not scale with the magnitude of the coefficient (e.g. when comparing the inter-die differences between triple-well on N-well junctions of Teddy design, devs. #7 and #8 (TED-7,8-3N)) whilst when they are large, noise that does scale with the fractional coefficient takes over (e.g. Teddy design, devs. #1 and #4 (TED-1,4)). This matter requires further study before it can be confirmed or disproved, but the key point is that the way in which noise affects power coefficients is not a trivial matter to uncover. Inter-die consistency merely allows a glimpse into the symptoms emanating therefrom.

- Negative power coefficients may be possible because in our results interpretation methodology the definitions for the extents of the areal and side-wall components of each junction are not sufficiently flexible (the matter is further discussed in section 6.5).

- The properties of pn-junctions are influenced by other doped species residing in their vicinity. A good example of that phenomenon is given by very dissimilar power coefficients extracted from N-well on substrate (NS) junctions originating from pure N-well on substrate (NS) devices as opposed to those originating from n-diffusion on triple well on N-well on substrate (n3NS) devices. The extent and spatial distribution of these 'extra' dopants in the vicinity of a test pn-junction are also likely to play a key role in determining the properties of the test junction. This phenomenon requires further study[10].

Finally, a conclusion that seems to generalise less well concerns the possibility that in some technologies fringe areas of certain diffusion-based junctions seem to contribute very little to the overall power output of the device at least when side-wall junction components are measured by means of the nominal perimeter of their host junction (Ninja design, devs. #4 and #9, N-well on substrate junctions (NIN-4,9) or Teddy design, devs. #2 and #3 (TED-2,3) for example).

6.4.3 Effects of junction type on power harvesting performance

The process of extracting power coefficients from various devices has already shown that: a) Different pn-junction types tend to behave in distinct ways and b) The presence of other dopants in the immediate vicinity of a pn-junction will affect its behaviour significantly. This implies that the immediate environment of a pn-junction is as much part of the junction type specification as the doped regions technically constituting the p- and n-sides of the said junction.

Another important consideration concerns the technological node and specific manufacturing process through which various structures are constructed. There is no reason to believe that different manufacturing technologies, more so if they represent different technological nodes, will provide the same absolute values of power coefficients or even the same quantitative relations between power coefficients relating

[10] A few possible effects that explain this situation are, however, known. For example, the presence of foreign doped species can affect carrier life-times and thus influence the efficiency by which electron-hole pairs are actually used for powering a load. The physics is not new, but a full-fledged, first-principle-derived study of the particular effects in CMOS technology lie outside the scope of this thesis. Note: this is a 'bulk' phenomenon that is not directly linked to the depletion region of the pn-junction. On the other hand, it is also known that the depletion region, on the other hand, will itself be influenced as its geometry is determined by diffusion patterns, which in turn depend on the interactions between dopant species.

to different junction types. For that reason we maintain the separation of power harvester designs by manufacturing technology throughout this section.

Data processing that will help understand how different junction types behave when tasked with power scavenging duties will involve comparing areal and side-wall power coefficients. This will involve 'plotting the behaviour' of each device grouping on a plane with the x-axis representing the areal and the y-axis the side-wall power coefficient value. Where more than one die is available, data points originating from the same device grouping on different dies will be plotted individually. Where groups of more than two devices are available it would be possible to plot data points corresponding to each possible pairing between devices belonging to the group. This will only be performed for 'interesting' pairings.

Unfortunately, there is too little information to allow for the determination of reliable confidence boundaries when it comes to the exact location of the power coefficients meant to represent each junction type. This problem is particularly acute when only a single pair is responsible for power coefficient extraction. Nevertheless, placing the power coefficient 'coordinates' on a Cartesian plane should still offer some insight into the relative strengths and weaknesses of different junction types when acting as optical power scavengers.

Ninja die # 1 (NIN1)

Ninja die #1 (NIN1) has offered a relatively poor collection of data points. Three data points are available, of which one, namely Ninja die #1, dev. #6 (NIN1-6) consists of extrapolated data and refers to a device where the passivation layer covering every other device on every other die is missing. Thus, with Ninja design, dev. #6's (NIN-6) data point excluded from any meaningful analysis we are left with one point corresponding to the pure N-well on substrate (NS) junction type (arising from the device pair Ninja die #1, devs. #1 and #5 (NIN1-1,5)) and one pair corresponding to the n-diffusion on substrate (nS) type (arising from device pair Ninja die #1, devs. #4 and #9 (NIN1-4,9)). The resulting plot is shown in Figure 6.8.

The power coefficient plot makes it obvious that the N-well on substrate (NS) type junction is a far more efficient power capture structure than the n-diffusion on substrate (nS) type at both areal and fringe components. This large difference can be seen because the distance between the data points is large compared to their individual distances from the origin (i.e. the 100% unresponsive case where neither areal nor side-wall components contribute anything to maximum power output).

Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2)

The Svejk (SVJ) design family only offers three data points per die, but two of them arise as a result of fits between four to five devices. Thus the power coefficient extracts obtained stand a better chance of showing reasonable representative values although, of course, the small size of basic cells used in their corresponding devices may be introducing deviations from the 'true' areal and side-wall power coefficient values. This was discussed in slightly more detail in 6.4.2. The power coefficient locus for Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) can be seen in Figure 6.9.

Figure 6.9 shows clearly that even with the limited number of dies used there are clear differences between different junction types i.e. the differences between types are large compared to both inter-die

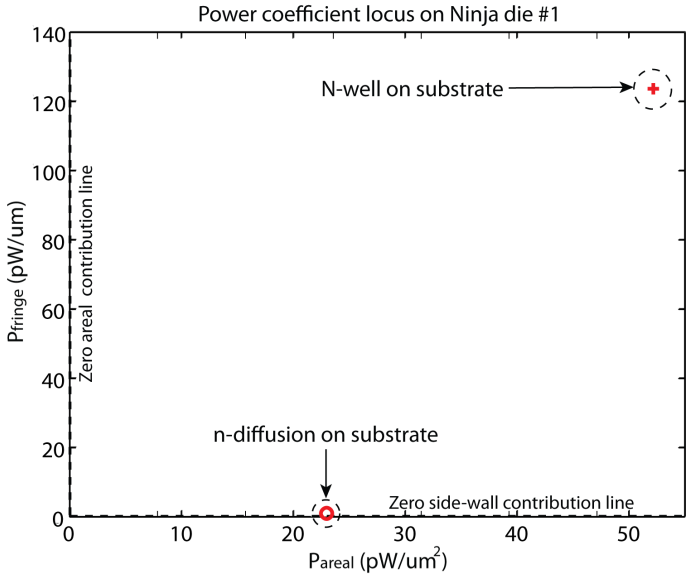

Figure 6.8: Normalised maximum power recovery per unit area or per unit perimeter (the 'power coefficients') locus for Ninja die #1 (NIN1). Each symbol represents a junction type. +: N-well on substrate (NS). o: n-diffusion on substrate (nS). Standard illumination was used (see table 6.3).

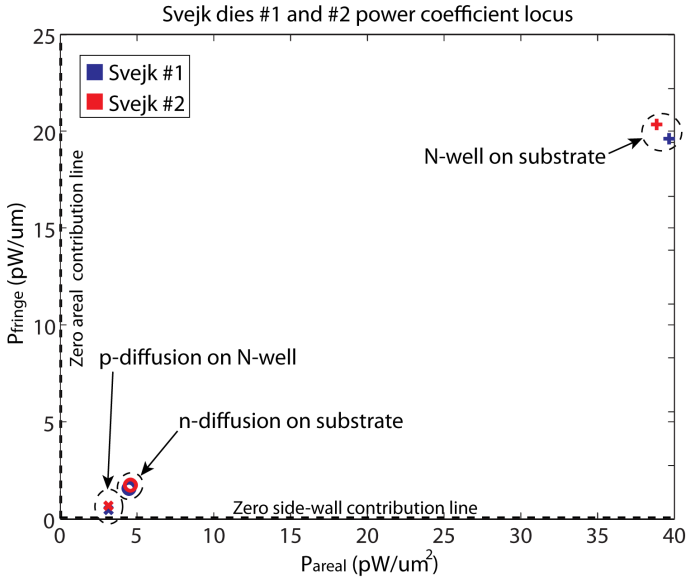

Figure 6.9: Normalised maximum power recovery per unit area or per unit perimeter (the 'power coefficients') locus for Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2). Each colour represents a die: Blue: Svejk die #1 (SVJ1). Red: Svejk die #2 (SVJ2). Each symbol represents a junction type. +: N-well on substrate (NS). o: n-diffusion on susbtrate (nS). x: p-diffusion on N-well (pN). Standard illumination was used (see table 6.3).

difference and to some extent also their distance from the origin. With more measured dies perhaps cluster analysis could have been pursued, but with just two available that would be superfluous. We notice how the N-well-based junctions (NS) show more inter-die variation than their diffusion-based counterparts, perhaps as a result of uncertainty factors that scale with the absolute value of the power coefficients.

Once again, the N-well on substrate (NS) junctions reveal themselves as the most responsive in terms of their areal component. On the other hand, diffusion-based junctions, although they appear to perform significantly differently, both share their propensity for low side-wall power coefficients and show a much smaller performance difference amongst themselves than when compared to their N-well on substrate (NS) counterparts.

In terms of relating these results to doping profiles we can state that the relatively high power coefficient values of the N-well on substrate (NS) junction are linked to the fact that typical CMOS technologies dope their N-wells more lightly than their diffusion regions and thus a well-on-substrate junction will effectively be a junction between the two most lightly doped regions anywhere on the die[11].

The same thought process might also be used to explain why n-diffusion on substrate (nS) junctions behave ever so slightly better than p-diffusion on N-well (pN) ones. If we assume the diffusion region doping profile on both n-diffusion on substrate (nS) and p-diffusion on N-well (pN) cases is similar, then the n-diffusion-based junctions will form a junction with the lightly doped substrate whilst the p-diffusion-based ones will form their junctions with the more heavily doped N-wells. This implies that the depletion region widths and also the photocurrent generation will be higher for n-diffusion on substrate junctions (nS). This hypothesis is based on certain, reasonable, but still to some extent speculative assumptions and therefore requires further study before it can be conclusively confirmed or disproved.

Teddy die #5 (TED5) and Teddy die #6 (TED6)

The Teddy (TED) design family has yielded a number of power coefficient results, all from device/junction pairings. We will split pn-junctions in to two categories: The first category (cat-1) will consist of all junctions that are formed in devices with the minimum number of dopant types necessary to sustain the junction type under consideration. This category includes n-diffusion on substrate (nS), pure N-well on substrate (NS), triple well on N-well (3N) in triple well on N-well on substrate (3NS), p-diffusion on N-well (pN) and n-diffusion triple well (n3) type junctions. The second category (cat-2) will include junctions that host superfluous dopant types in their ranks such as N-well on substrate (NS) in triple well on N-well on substrate (3NS) (unnecessary triple well (3W) region), N-well on substrate (NS) in n-diffusion on triple well on N-well on substrate (n3NS) (unnecessary n-diffusion (n+) and triple well (3W) regions) etc. All power coefficients involving cat-1 junctions (bar the outliers we have removed in section 6.4.2) are summarised in Figure 6.10. Notably, inter-die consistency seems reasonably good for all junctions.

Interestingly, N-well on substrate (NS) type junctions prove the most responsive with both areal and side-wall power coefficients standing clear above those of all of their competitors. This is not surprising since N-well on substrate (NS) junctions are expected to be the type formed between the two most lightly

[11] Here we only consider the 'main' doped regions consisting of p-diffusion, n-diffusion, triple-well, N-well and the substrate. A study of more elaborate regions such as lightly doped drain (LDD) regions in advanced MOSFETs lies outside the mandate of this thesis.

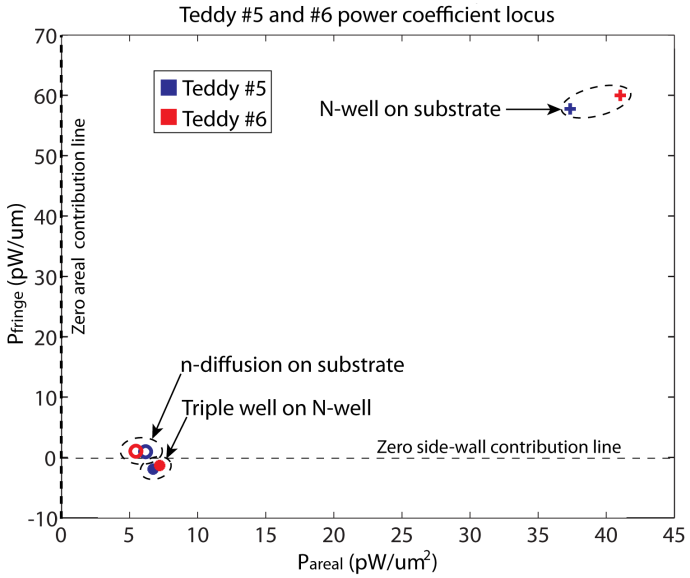

Figure 6.10: Normalised maximum power recovery per unit area or per unit perimeter (the 'power coefficients') locus for 'interesting' junctions of Teddy die #5 (TED5) and Teddy die #6 (TED6). Each colour represents a die: Blue: Teddy die #5 (TED5). Red: Teddy die #6 (TED6). Each symbol represents a junction type. +: pure N-well on substrate (NS). *: triple well on N-well (3N) in N-well on substrate (3NS) o: n-diffusion on substrate (nS). x: n-diffusion on triple well (n3). Standard illumination was used (see table 6.3).

doped regions available for CMOS manufacturing (excluding specialised regions such as LDD etc.).

On the other hand it seems that both other 'pure', cat-1 junctions show generally low fringe coefficients. For diffusion-based junctions we have already explained why that might be the case throughout this section (resultprocpower), but triple well on N-well (3N) junctions were expected to feature a significant 'vertical' junction component[12] that would contribute to the overall power output by a sizable amount. The exact reasons for such low fringe coefficient in triple well on N-well (3N) structures is therefore as yet unknown.

Cat-2 junctions ('contaminated' ones) exhibit very erratic behaviour characterised by some exceptionally high power coefficients (with some fringe power coefficients in the nW per unit area or length or one hundred to one thousand times more powerful than most other such coefficients in any technology and for any junction type). As such, studying them is rendered futile by the fact that the structures we have designed are simply not designed in a way that allows meaningful analysis of data related to such complicated junctions. Moreover, we have already established that the behaviour of cat-2 junctions is a function of not just their own geometrical data, but also the geometrical data of the other dopants that they share their vicinity with. For that reason we shall not attempt such analysis.

Nevertheless we shall once again note the interesting behaviour of triple well on N-well (3N) devices in n-diffusion on triple well on N-well on substrate (n3NS) junctions, for example Teddy design, devs. #7 and #8 (TED-7,8): these devices showed surprisingly consistent power coefficients (low inter-die variation) and 'reasonable' absolute values for them (no negative values, no exceptionally large values etc.). This may be an indication that certain junctions can be relatively shielded from the distribution of

[12] Vertical junction component: the portion of the metallurgical surface that lies approx. perpendicular to the surface of the die.

certain additional dopant species in their vicinity. For example, if we take the results for Teddy design, devs. #7 and #8 (TED-7,8) at face value we can deduce that the distribution of n+ regions near triple well on N-well (3N) junctions does not dramatically affect the function of the triple well on N-well (3N) junction. However, the presence or absence of the aforementioned doped species does make a difference. At the same time both triple well on N-well (3N) and N-well on substrate (NS) components of Teddy design, devs. #7 and #8 (TED-7,8) and Teddy design, devs. #5 and #9 (TED-5,9) behaved very differently between them. This then leads to one more possible inference: n+ regions in triple well on N-well (3N) affect the function of both triple well on N-well (3N) and N-well on substrate (NS) junctions mostly by their general presence. The exact layout pattern doesn't matter that much as the influence is exerted in a geometrically very coarse pattern. Once again, uncovering the full complexity of effects would require further study.

Conclusions

We can visualise all 'intersting' power coefficients (i.e. those that we have reasons to believe have yielded meaningful results) in Figure 6.11. We notice how the N-well on substrate type junctions offer good areal and side-wall contributions to power recovery at the maximum power point with all other junctions struggling to obtain any sizeable contribution from their side-wall components. Diffusion-based and triple-well on N-well (3N) junctions also show similar, although apparently distinct performances, most notably in terms of areal power coefficients. It is only the C35 die that appears to feature n-diffusion on substrate (nS) junctions with significantly better performance than similar junctions in other technologies, but the effects of technology on power delivery will be examined in more detail farther down this chapter.

In summary, studying power coefficients with an eye towards understanding how the junction type affects power scavenger performance we have reached the following conclusions:

- N-well on substrate (NS) type junctions generally tend to be the most reliable and responsive both in terms of their areal component contribution and in terms of their side-wall component contribution to overall power output.
- p-diffusion on N-well (pN) and n-diffusion on substrate (nS) type junctions did show differences amongst themselves, but apart from explanations revolving around the differences in doping profiles seen at those junctions we could possibly attribute some of the difference to uncertainties introduced by e.g the small size of basic cells and the fact that areal and side-wall junction components are not that clearly separable. Thus, the power coefficient values obtained for each junction type cannot necessarily be taken at face value.
- Junctions can be separated in cat-1 and cat-2 'flavours'. Cat-1 junctions are constructed of the minimum number of species of dopants required to implement them whilst cat-2 junctions feature other, not strictly necessary species of dopants in their vicinity. Cat-2 junctions exhibit far more complicated power harvester performance behaviour which is a function of not just the geometrical features of the doped regions used to create them, but also the geometrical features of the additional doped regions in the vicinity.
- Inter-die differences for cat-1 junctions seem to be generally small compared to differences in per-

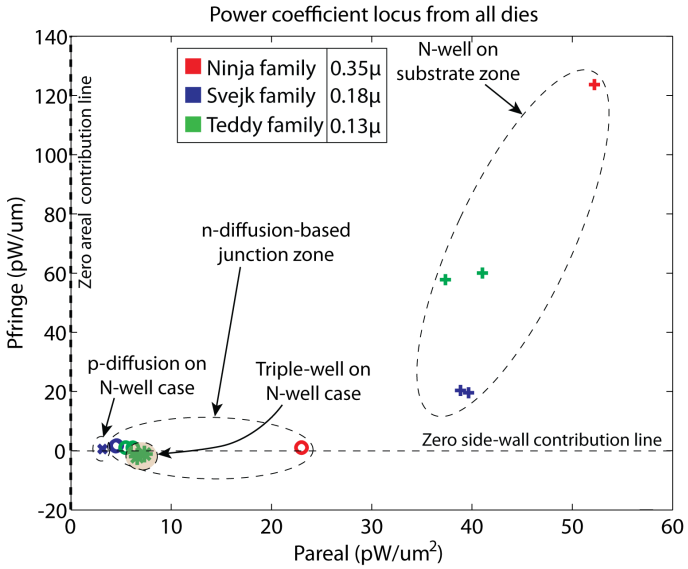

Figure 6.11: Normalised maximum power recovery per unit area or per unit perimeter (the 'power coefficients') locus for 'interesting' junctions from all our test dies. Each colour represents a die design family (and consequently also a technological node: Red: Ninja - 0.35 micron. Blue: Svejk - 0.18 micron. Green: Teddy - 0.13 micron. Each symbol represents a junction type. +: pure N-well on substrate (NS). *: triple well on N-well (3N) in N-well on substrate (in 3NS) o: n-diffusion on substrate (nS). x: p-diffusion on N-well (pN). Standard illumination was used (see table 6.3). Note: the blue 'x' and 'o' marks actually represent two points that are so close to each other, they become impossible to distinguish in this plot.

formance between different types of junctions. Inter-die discrepancies may tend to be exacerbated for weak power harvesters because I-V sweeps obtained from such devices is more vulnerable to noise that does not scale with photocurrent magnitude.

- In general we see some evidence that power harvester performance is directly related to our expectations regarding doping profiles at each junction type. Thus, N-well on substrate (NS) type junctions are expected to feature the widest depletion regions with diffusion-based junctions featuring the narrowest. Lack of specific doping profile information, however, means that linking obtained results to doping profiles is largely speculative, yet based on related information (e.g. I-V sweeps from [3]).

6.4.4 Effects of technological node on power harvesting performance

In order to understand how the choice of technological node affects power harvester performance we need to compare power coefficients extracted for the same type of pn-junctions (ideally cat-1) on dies manufactured in different technologies. In the library of our optical device designs the set of all junction types that occur in more than one technology consists of just the N-well on substrate (NS) and n-diffusion on substrate (nS) junction types. Both these types appear in all three technologies of choice. We can therefore pinpoint N-well on substrate (NS) and n-diffusion on substrate (nS) areal and fringe power coefficients from each die on a Cartesian plane. The resulting plot is shown in Figure 6.12. Note: Ninja

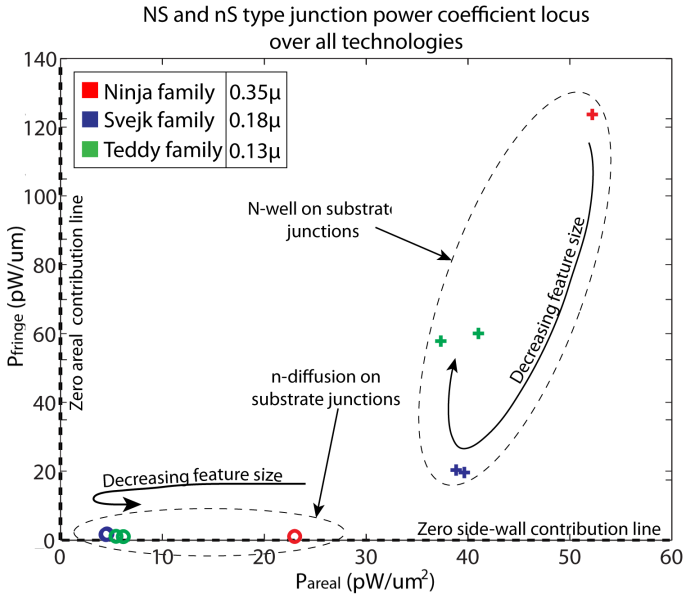

Figure 6.12: Normalised maximum power recovery per unit area or per unit perimeter (the 'power coefficients') locus of N-well on substrate (NS) and n-diffusion on substrate (nS) type junctions over all dies. Each colour represents a CMOS manufacturing technology. Red: 0.35 micron AMS35. Blue: 0.18 micron IBM18. Green: 0.13 micron UMC13. Each symbol represents a different junction type. +: N-well on substrate (NS). o: n-diffusion on substrate (nS). Standard illumination was used (see table 6.3). Note: the blue circles represent two data points that are too close to each other to distinguish in this plot.

die #1 (NIN1) will only contribute those N-well on substrate (NS) power coefficients that correspond to devices with the passivation cover on top of them.

We can see from the figure that whilst indeed the AMS35 die Ninja die #1 (NIN1) did show the most photo-sensitive N-well on substrate (NS) junctions and at least the most sensitive areal n-diffusion on substrate (nS) junctions, the trend does not continue smoothly between the IBM18 and the UMC13 dies. Indeed UMC13 dies seem generally more sensitive to light than their IBM18 counterparts despite the fact that one would expect higher doping concentrations and narrower depletion regions in the 0.13 micron technology. This could be as a result of the specific front-end manufacturing techniques used in either case (the IBM18 and UMC13 technologies are manufactured by different foundries, which makes this possibility distinctly more likely).

However, there is one more possible explanation for the unexpected break in the inter-technology trend. We noticed that technology H18 is coated by a thick polyimide layer that sits atop the standard Silicon dioxide (SiO_2) and Silicon Nitride (Si_3N_4) passivation layers. Thus, when light enters our C35 and UMC13 dies it has to go through Si_3N_4 first, then through SiO_2 and then any parts of the illumination not blocked by metallisation, proceeds to the front-end side, whilst in H18 the polyimide forms the first 'barrier'.

This hypothesis merits a back-of-the-envelope calculation. The analysis begins by the following observations and simplifying assumptions:

- If the polyimide layer is assumed to be a few microns thick and its optical transmittance at visible

wavelengths is good, then optical losses through the polyimide can be ignored.

- If the point above holds, then we only need to consider reflective losses at the air-polyimide interface and at the polyimide-Si_3N_4 interface and compare them with the air-Si_3N_4 loss that would occur in the absence of the polyimide.
- In order to keep the calculations simple, we shall assume that all inter-material interfaces are planar, parallel to each other and perpendicular to the incoming light.
- We shall only consider the 'main' illumination component that gets transmitted through each interface. Thus, the illumination component passing from the air to the polyimide, then getting reflected at the polyimide-Si_3N_4 interface, then getting internally reflected at the polyimide-air interface and then bounding back and ultimately entering the Si_3N_4 will not be considered.
- None of our dies has been manufactured with an anti-reflective coating. This has occurred because no such layers were available at the time of manufacture in technology C35 and as such for the sake of comparison dies manufactured in other technologies lacked it as well.

Important parameters in this problem include:

- The refractive index of air is taken as 1 for all wavelengths.
- Si_3N_4 refractive index: approx. 2 for the $600 - 1000\,nm$ wavelength range, between 2 and 2.1 for the $400 - 600\,nm$ wavelength range. Here we simply use the value of 2.
- Polyimide refractive index: Polyimides being a group of materials it is difficult to find any reliable data pertaining to their properties and even if we did, we cannot be certain that the polyimide coating on H18 dies is the same as the one whose properties we would be quoting. Finally, polyimides are polar molecules and that change their refractive index as a function of frequency when in solution. However when encountered as solids we have no reason to expect that their refractive indices will be different to that at very high frequencies (high enough so that molecules cannot arrange themselves in the lowest energy configuration). Finding no reliable refractive index vs. wavelength plots we use data in [4, 5] and simply assume a 'reasonable' refractive index of 1.8 for our calculations.
- Polyimide transmittances as a function of wavelength can be found in [5][13]. Given the transmittance spectra given by the author show transmittance as always over 70% at $600\,nm$ wavelength and at least up to $1200\,nm$, and these refer to films of 40 microns thicknesses, we conclude that our assumption that a polyimide layer of a few microns width should not significantly hamper optical transmission is acceptable.

We may now proceed with the computation of optical losses/transmittance from the air to the Silicon nitride for the main illumination component in both cases. We shall use the perpendicular incidence Fresnel equations to compute transmittance:

[13] Absorbances as functions of wavelength can also be found in [4] for Ti-bearing polyimides.

$$T = 1 - \left| \frac{n_i - n_t}{n_i + n_t} \right|^2 \tag{6.2}$$

where T is the transmittance coefficient, n_i the 'incidence-', 'source-' side refractive index and n_t the 'transmittance-' or 'destination-' side refractive index. The equation holds for light of any polarisation.

In the case where no polyimide coating is applied to the die we only suffer losses related to the air-Si_3N_4 interface. Thus, the transmittance can be described by the air-Si_3N_4 transmittance coefficient. The value yielded is approx. 88.9% transmission.

In the case where there is a thin polyimide coating on top of the Si_3N_4, we will suffer losses first at the air-polyimide interface, and then at the polyimide-Si_3N_4 interface. Thus the overall main optical component transmittance can be described by the product of the transmittance coefficients for both interfaces. The air-polyimide interface shows a transmittance of approx. 91.8%, while the polyimide-Si_3N_4 interface shows transmittance of 99.7%. The total transmittance is therefore approx. 91.5% for the case with the polyimide coating.

We thus notice how with the parameter values, simplifications and assumptions we made, it would seem that the polyimide coating makes only a marginal difference to the overall transmittance of visible light through the back-end of each die. Therefore, either our model/computations were inaccurate, or the discrepancies observed in terms of normalised maximum power recoveries per unit area and length are not due to the extra polyimide coating in technology H18 and have to be attributed to differences in the front-end.

Note: this is only a back-of-the-envelope calculation. In reality not only the numbers may differ, but we expect some differences in layer thicknesses and exact composition between different manufacturing technologies. We do not have access to specific data from each technology so comparing the effects of back-end layers on transmittance between technologies was not carried out. Another reason for this omission was that the issue was only considered at a late stage in the project as it was up to that point assumed that all technologies used very similar layer structures and thicknesses. The very similar performances of Ninja design, devices #5 and #6 (NIN-5,6) despite the absence of the entire passivation layer hint towards the possibility that our assumption was correct (see table 6.4). However, further study is required before our assumption can be either confirmed as true or rejected as provably wrong.

Finally, we observe that power coefficients change along more or less straight lines for the different technologies. This would indicate that despite differences in the specifics of their manufacturing techniques, areal and side-wall components of pn-junctions tend to change their properties in similar fashions from (technological) node to node.

6.4.5 Power harvesting designs under the constraints of layout

When designing a real power scavenging element for CMOS integration we will typically strive to obtain the maximum amount of power out of specified illumination conditions under the constraints of a limited area within which the element needs to fit. Thus, design for manufacturing will be significantly different to design for academic research. In this small subsection we shall mention a few considerations that come into play when translating power harvester designs into a more practical setting and consider some of the data gathered throughout this chapter under a somewhat different light.

To begin with, it might be worth normalising the maximum power output provided by each designed device to its footprint and determine how much power can be obtained per square micron of footprint. Naturally the specifics of the layout, including metallisation coverage, total areal and side-wall extents, specific shape of the overall junction etc., will influence the final output, which is the reason why this exercise should not be considered as a complete, exhaustive description of the capabilities of each junction type in each technology. It will, however, provide an opportunity to compare junctions that either haven't been analysed because there were no available devices for pairing or have been analysed but the results were inconclusive (for example in junctions whose properties have been altered by the presence of extra dopant species in the vicinity).

Furthermore, it would be useful to generate a 'good design handbook' that describes how each junction should be laid out for optimal power recovery capability within a limited layout area. Unfortunately, this is no trivial task due to not only the multitude of design rules and power coefficients that relate to various junction types, but also because of deviations from the simplistic 'areal plus side-wall' junction model at small basic cell sizes and the presence of metallisation in order to ensure good electrical contact to all parts of each doped region. Other considerations, such as how thin or thick the metallisation is compared to the minimum design width of a rectangular doped region of some type or the doped region's sheet conductivity will also affect layout strategies. Therefore we leave the determination of optimum layout strategies as a topic for future work (the issue is simply too design-rule specific to warrant the determination of generic, optimal geometric patterns).

Note: like all extracted power coefficients, any optimal design strategy will be dependent on the spectral content of the illumination falling upon the junction under consideration. It is a well-known phenomenon that pn-junctions with areal components at different depths from the surface of the die will respond optimally to different wavelengths. This is another reason why the determination of optimum layout strategies should be relegated to computer programs, working with parametric data provided by detailed, dedicated studies on the subject, ideally provided by the technology manufacturer.

Finally, we must mention that whilst this section has been concerning itself with maximum power delivery from various junctions, the voltage at which this power is available is also very important because it determines how many transistors of given threshold voltage we can 'stack' within the headroom yielded by the power scavenger cells. A quick look through the relevant data (tables 6.4, 6.5, 6.6, 6.7 and 6.8) shows that N-well on substrate (NS) devices once again prove to be the best. Unfortunately, because N-well on substrate (NS) devices have the substrate (i.e. GND) as their positive terminal they are unsuitable for powering circuitry that resides on the same die. The same applies for n-diffusion on substrate (nS) junctions too. This only leaves the p-diffusion on N-well (pN), triple well on N-well (3N) and n-diffusion on triple well (n3) types as valid candidates for the role of solar batteries. According to the same tables, they all provide approx. $0.4 - 0.5\,V$ at open circuit.

In conclusion, in the following sections we have summarised key data from each device on Ninja die #1 (NIN1) (6.22), Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) (6.23) and Teedy die #5 (TED5) and Teddy die #6 (TED6) (6.24). This includes measured maximum power output, total footprint size, maximum power output per unit footprint area and some useful geometrical data, i.e. areal and side-wall 'junction densities' (hence-forth also called 'j-densities'). These metrics are a way of assessing how 'densely' each square micron of footprint area is populated by areal and side-wall junctions respectively. The units for these metrics will be: a) For areal junctions square microns of areal junction per square

micron of footprint (i.e. dimensionless. Symbol: D_A.). b) for side-wall junctions microns of perimeter length per square micron of footprint area (i.e. units of $1/\mu m$. Symbol: D_P.).

6.4.5.1 The Ninja family performance

We begin by examining Ninja die #1 (NIN1). The table is sorted in descending order with respect to the power per unit footprint area (P/F) metric. Ninja die #1, dev. #6 (NIN1-6) was omitted due to its special design in terms of the passivation layer. Figure 6.13 helps visualise the table.

Table 6.22: Power density (P/F) for all devices on Ninja die #1 (NIN1) (defined as maximum power output per unit footprint area). Standard illumination was used (see table 6.3).

POWER DENSITY: NIN						
DEV	**TYPE**	P_{max}	**Footprint**	**P/F**	D_A	D_P
		μW	μm^2	$\mu W/mm^2$	$\mu m^2/\mu m^2$	$\mu m/\mu m^2$
1	NS	11.6300	229441	50.69	0.33	0.27
5	NS	11.5114	229441	50.17	0.58	0.16
6	NS	10.8832	229441	47.43	0.58	0.16
2	NS	9.1029	229441	39.67	0.12	0.66
4	nS	4.7648	229441	20.77	0.90	0.05
9	nS	2.1259	229441	9.27	0.34	1.33
2	pN	1.3631	229441	5.94	0.37	0.05

Shown is also other relevant information such as the actual maximum power recorded, the footprint area and the densities of areal (D_A) and side-wall (D_P) junctions per unit footprint area for each device under consideration.

We observe that the top spots are all taken by N-well on substrate (NS) type junctions. Interestingly, the top spot is taken by a design with balanced distribution of areal and side-wall components whilst the second spot is occupied by an overwhelmingly areal-dominated[14] junction and the third spot by an overwhelmingly side-wall dominated design. This result cannot be taken at face value, but perhaps indicates that in the AMS35 technology there may be no straightforward optimal solution to the issue of the optimum layout technology and certainly places a lower bound as to what is achievable (at least under our specific source of illumination).

Diffusion-based junctions trail behind with n-diffusion on substrate (nS) beating p-diffusion on N-well (pN) by a large margin. The front-runner in the n-diffusion on substrate (nS) category is overwhelmingly areal-dominated which hints towards the possibility that in C35 designing large slabs of continuous n-diffusion on substrate (nS) junctions can possibly yield the highest power recovery densities for that type of junctions. Moreover, such simple, areal-dominated designs can be serviced by relatively simple and rare networks of metallisation and associated contacts. This is because of the absence of complicated shapes and current bottlenecks.

Note: the p-diffusion on N-well (pN) device considered has significantly lower areal junction density compared to the areal-dominated n-diffusion on substrate (nS) device #4 competitor (NIN1-4) at 37 % vs. 90 % coverage, but the employed layout is not necessarily the one that can yield the optimal results

[14] Admittedly the notion of areal-dominated and side-wall dominated junctions on the mere basis of junction densities is somewhat arbitrary, but we shall nevertheless use it as a practical and 'user-friendly' means of describing the relative 'weights' of each junction component in the whole.

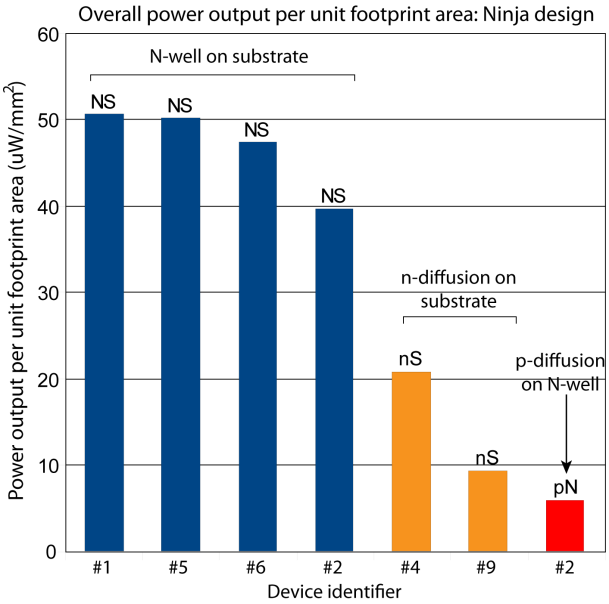

Figure 6.13: Maximum power outputs per unit footprint area for devices residing on the 'Ninja' die. Letter pairs on top of each bar represent the sub-junction under test within each device. Standard illumination was used (see table 6.3).

for this junction. Nevertheless, the relatively large difference between n-diffusion on substrate (nS) and p-diffusion on N-well (pN) performance still hints towards the fact that given the choice, n-diffusion on substrate (nS) junctions should be preferred in technology C35.

6.4.5.2 The Svejk family performance

Next, we consider the Svejk (SVJ) design family. Key metrics, including power density, are summarised in table 6.23. Because two dies were available, we have pooled the maximum delivered power for each device between dies by taking the average. Results are sorted in descending order with respect to power density. Table 6.23 can be visualised with the aid of Figure 6.14.

Data shows that in technology H18, much like in C35, NS devices are solidly holding all the top spots. In terms of areal and side-wall contributions we can say that there is no obvious trend in the data. The first and third devices, for example, have fairly balanced areal and side-wall components, easily outperforming counterparts with much more important side-wall components (e.g. Svejk design, dev. #8 (SVJ-8) and Svejk design, dev. #10 (SVJ-10)) or devices with much more dominant areal components (e.g. Svejk design, dev. #3 (SVJ-3) or Svejk design, dev. #4 (SVJ-4)). Notably, for Svejk design, devs. #8, #9, #10 and #11 (SVJ-8,9,10,11) the sum of areal and side-wall components remains roughly the same, which means that by changing basic cell size as we did during the design phase of these devices we have effectively exchanged 1 square micron of areal component for roughly 1 micron of side-wall. This is interesting given that by using this method we have found an easy way of 'gearing' finite footprint photoelectric element design either into overwhelmingly areal mode or into overwhelmingly side-wall mode. This strategy is most likely not optimal, but very easy to implement.

Next, we find that the triple well on N-well (3N) device, for which there is no pairing partner

Table 6.23: Power density (P/F) for all devices on Svejk die #1 (SVJ1) and Svejk die #2 (SVJ2) (defined as maximum power output per unit footprint area). Standard illumination was used (see table 6.3).

POWER DENSITY: SVJ						
DEV	**TYPE**	$\overline{P_{max}}$	**Footprint**	**P/F**	D_A	D_P
		μW	μm^2	$\mu W/mm^2$	$\mu m^2/\mu m^2$	$\mu m/\mu m^2$
11	NS	1.1098	40000	27.75	0.61	0.24
12	NS	1.0669	40000	26.67	0.61	0.10
9	NS	1.0182	40000	25.46	0.40	0.41
8	NS	0.9292	40000	23.23	0.27	0.58
10	NS	0.9224	40000	23.06	0.27	0.58
7	NS	0.6044	40000	15.11	0.18	0.56
3	NS	0.8451	90000	9.39	0.35	0.00
4	NS	0.5081	90000	5.65	0.08	0.00
5	nS	0.2158	90000	2.40	0.43	0.28
12	3N	0.0482	40000	1.21	0.35	0.58
6	nS	0.1062	90000	1.18	0.24	0.00
3	pN	0.1058	90000	1.18	0.34	0.19
2	nS	0.0497	90000	0.55	0.13	0.00
1	nS	0.0425	90000	0.47	0.13	0.00
4	pN	0.0213	90000	0.24	0.08	0.00

Shown is other relevant information such as the actual maximum power recorded, the footprint area and the densities of areal (D_A) and side-wall (D_P) junctions per unit footprint area for each device under consideration.

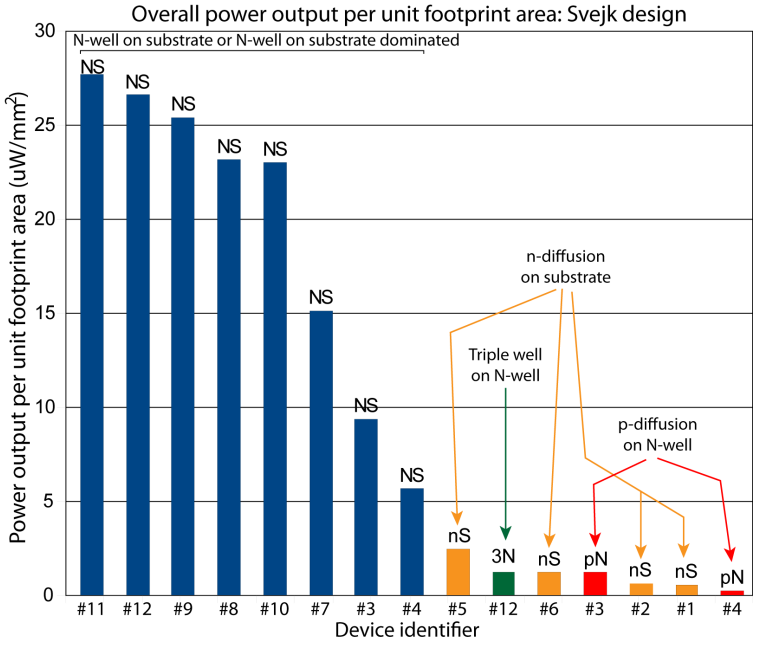

Figure 6.14: Maximum power outputs per unit footprint area for devices residing on 'Svejk' type dies. Data from homologous devices residing on different dies has been pooled so that for each device identifier only the average maximum power output is given. Letter pairs on top of each bar represent the sub-junction under test within each device. Standard illumination was used (see table 6.3). Note: purple bars indicate devices tested with all their pn-sub-junctions in parallel (e.g. p-diffusion on N-well (pN) in parallel to N-well on substrate (NS)). These devices are evidently dominated by the performance of the N-well.

available on Svejk (SVJ) devices, is showing similar overall performance to its diffusion-based competitors. However, when compared to Svejk design, device #3 (SVJ-3), a p-diffusion on N-well design with the same areal coverage but smaller fringe component, the triple-well on N-well device (3N) fails to capitalise on its extra side-wall junctions. This is a hint towards the possibility that triple well on N-well (3N) junctions are practically inferior to p-diffusion on N-well (pN) designs, especially since diffusion regions can typically be laid out with smaller feature sizes than wells of all types, which makes the p-diffusion on N-well (pN) type slightly more versatile. This means that if the layout design is really pushed to its limits, it is likely that p-diffusion on N-well (pN) devices can be manufactured with higher junction densities than triple well on N-well (3N) devices and thus possibly offer clearly higher power densities in the end.

Trailing at the end of the table are all the n-diffusion on substrate (nS) and p-diffusion on N-well (pN) junctions. We notice how the differentiations between these devices now seem to hinge primarily on their junction densities rather than the type of junction under consideration.

6.4.5.3 The Teddy family performance

Finally, we visit the Teddy (TED) design family. Key metrics sorted in a fashion similar to the corresponding tables for the Svejk (SVJ) and Ninja (NIN) design families are shown in table 6.24. Because two Teddy (TED) type devices were available, we have pooled maximum delivered power for each device between the two dies by taking the average. the results of table 6.24 can be visualised in Figure 6.15.

Table 6.24: Power density (P/F) for all devices on Teddy die #5 (TED5) and Teddy die #6 (TED6) (defined as maximum power output per unit footprint area). Standard illumination was used (see table 6.3).

POWER DENSITY: TED						
DEV	**TYPE**	$\overline{P_{max}}$	**Footprint**	**P/F**	D_A	D_P
		μW	μm^2	$\mu W/mm^2$	$\mu m^2/\mu m^2$	$\mu m/\mu m^2$
8	NS	4.0929	89401	45.78	0.82	0.09
9	NS	3.6018	89401	40.29	0.84	0.09
1	NS	9.2129	245025	37.60	0.14	0.54
4	NS	9.2066	245025	37.57	0.89	0.05
7	NS	3.1352	89401	35.07	0.81	0.09
5	NS	7.8880	245025	32.19	0.75	0.07
2	nS	1.1957	245025	4.88	0.52	1.82
3	nS	1.1838	245025	4.83	0.81	0.09
9	3N	0.2911	89401	3.26	0.53	0.27
8	3N	0.2541	89401	2.84	0.52	0.27
7	3N	0.2504	89401	2.80	0.70	0.09
5	3N	0.4229	245025	1.73	0.18	0.53
7	n3	0.1147	89401	1.28	0.59	0.09
8	n3	0.0509	89401	0.57	0.22	0.18

Shown are other relevant information such as the actual maximum power recorded, the footprint area and the densities of areal (D_A) and side-wall (D_P) junctions per unit footprint area for each device under consideration.

The Teddy (TED) dies, just like their Svejk (SVJ) and Ninja (NIN) counterparts, show very clear

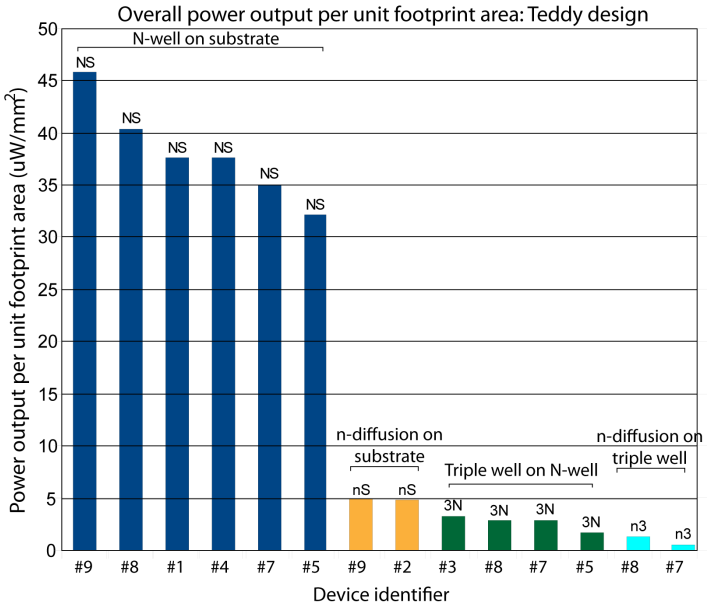

Figure 6.15: Maximum power outputs per unit footprint area for devices residing on 'Teddy' type dies. Data from homologous devices residing on different dies has been pooled so that for each device identifier only the average maximum power output is given. Letter pairs on top of each bar represent the sub-junction under test within each device. Standard illumination was used (see table 6.3).

trends: N-well on substrate (NS) devices dominate the top in a solid block. n-diffusion on substrate (nS) devices follow with triple well on N-well (3N) devices falling behind n-diffusion on substrate (nS) and n-diffusion on triple well (n3) devices trailing at the end.

Amongst N-well on substrate (NS) devices we see the top devices (both areal-dominated) being very similar in structure and yielding similar results. Given that both Teddy design, dev. #8 (TED-8) and Teddy design, dev. #9 (TED-9) have very similar N-well on substrate (NS) and triple well on N-well (3N) basic cells (see corresponding entries in table 6.1) and only differ between themselves by the presence of n-diffusion in Teddy design, dev. #8 (TED-8) and its absence in Teddy design, dev. #9 (TED-9) we can infer that exactly this difference is the most likely cause of the performance discrepancy between Teddy design, dev. #8 (TED-8) and Teddy design, dev. #9 (TED-9). The presence of n-diffusion could be affecting performance by altering the doping profile of the underlying structures, but this is a matter for further studies at semiconductor level. Note: The shape, size and distribution of both N-well on substrate (NS) and triple well on N-well (3N) basic cells on Teddy design, dev. #8 (TED-8) and Teddy design, dev. # (TED-9) is identical, so differences in the distribution of triple well (3W) regions over the two devices are unlikely to be the causes of the performance discrepancy.

The above-stated hypothesis, however is not needed to explain the performance discrepancies between the N-well on substrate (NS) components of Teddy design, dev. #7 (TED-7) and Teddy design, dev. #5 (TED-5): We notice that the ratio of junction densities is approximately equal to the ratio of power densities (1.08 vs. 1.09), hence the performance difference can be almost entirely explained by the difference in areal junction densities (the side-wall junction densities are almost identical and very low). Caveat: the distribution of triple well (3W) in Teddy design, dev. #5 (TED-5) and Teddy design, dev. #7 (TED-7) is dramatically different and this might conceivably be canceling out the effects of the presence/absence of n-diffusion (it would be an extraordinary coincidence if it did but the possibility can

not be excluded).

In terms of triple well on N-well (3N) junctions we observe a very obvious performance discrepancy between Teddy design, dev. #8 (TED-8) and Teddy design, dev. #9 (TED-9) (similar to their N-well on substrate (NS) counterparts). Once again junctions with very similar j-densities behave very differently. This too can be potentially attributed to the effects of the presence of n-diffusion on one device, but not on the other. Because of the highly dissimilar j-densities of the triple well on N-well (3N) components of Teddy design, dev. #5 (TED-5) and Teddy design, dev. #7 (TED-7) we can, sadly, not easily confirm that this effect takes place for that pair of devices too.

n-diffusion on substrate (nS) junctions surprise us by being reasonably capable of competing with triple well on N-well (3N) contestants. Notably, n-diffusion on substrate (nS) type device Teddy design, dev. #3 (TED-3) and the triple well on N-well (3N) component of device Teddy design, dev. #7 (TED-7) feature similar junction densities, but the n-diffusion on substrate (nS) junction features significantly higher power density. This indicates that in a practical setting the diffusion-based junction will perform better. Furthermore, the n-diffusion on substrate (nS) junction being diffusion-based is subject to less harsh layout rules and can be therefore laid out with more versatility that its well-based competitor.

Finally, n-diffusion on triple well (n3) devices trail the league far behind all of their other competitors. Sadly, no n-diffusion on triple well (n3) devices featuring high densities of both areal and side-wall components were manufactured.

As an overall conclusion we can therefore state that dense N-well on substrate (NS) type junctions generally tend to be the most preferable for the implementation of optical power harvesters under a source of illumination as described in 6.2.1. As for other junction types, tests need to be performed each time as it seems that the specifics of each technology make it difficult to correctly guess what junction types will perform best. The junction type ranking tables generally show different orders for each technology type.

6.4.5.4 Comparison with typical, commercial solar cell

In order to put our results into perspective we shall now compare the maximum power output per unit area of our test devices under standard illumination to the power output of a typical commercial solar panel under sunlight. We have randomly chosen the Panasonic VBHN220AA01 [15] with a stated energy conversion efficiency of 19.8 %. If we combine this information with the average July day irradiance in the Isle of Wight quoted before (equal to $463\,pW/\mu m^2$) we obtain an average daily electrical power output density of approx. $91.7\,pW/\mu m^2$ or equivalently $91.7\,\mu W/mm^2$.

This value bests the maximum electrical power outputs densities of every single test device we have used by a large margin (closest competitor: Ninja design, device #1 (NIN-1), an N-well on substrate (NS) device that can manage $50.69\,\mu W/mm^2$). However, it is noticeable that even though Ninja device #1 (NIN-1) only manages to produce slightly over half of the power density recorded by our sample solar cell, the comparison is with an average day in July, which is not expected to be far from the best a solar panel can achieve in the United Kingdom.

[15] Data-sheet found at http:// us.sanyo.com/ dynamic/ product/ Downloads/ Panasonic%20HIT %20220A%20 Data%20 Sheet_ WEB-10609680.pdf

6.5 Extending the linear fit model

Our linear fit model, described by equation 6.1 on page 259, offered some insight into the inner workings of real CMOS power harvesting elements, but after the experiments summarised in table 6.18 on page 263 we did notice that pairing devices with different basic cell sizes led to the extraction of different power coefficient values. As a result it was decided that the model should be extended with a new parameter 'd' that will denote the effective range within which the presence of a side-wall exerts influence over the performance of what has hitherto been considered to be areal junction territory.

In this section we shall first describe exactly how the model was extended before we proceed to showcasing the results stemming from the application of the extended model on a set of strategically selected test pn-junctions. Next, we attempt to explain the interaction of the model extension with the original 'areal + fringe' junction models by examining the case of negative side-wall coefficients recorded for some device pairings on our Svejk die family (see table 6.18) before we close the section with a discussion of the implications arising from our results.

6.5.1 Implementation of model extension

The philosophy behind the extension relies upon the observation that doping profiles in real CMOS junctions are not clear-cut, with areal and side-wall junctions intersecting each other at perfectly right angles and featuring perfectly homogeneous properties throughout their entire extents. Instead, even simple doping profiles will show smooth transitions between areal and side-wall regions, as we describe in section 3.3. This is what gives rise to the concept that within a certain range from a side-wall, areal junctions will start behaving 'significantly differently' from how they behave in the ideal case of an areal junction of infinite extent; where the exact meaning of 'significantly' is determined by the application and the accuracy it requires.

Thus, instead of splitting pn-junctions into areal and side-wall components we shall now be splitting them into 'core area' and 'border regions', where the latter are defined as any area lying within a distance of at most 'd' from their nearest side-wall edge. The end effect is that once we introduce the 'd' parameter we precipitate a qualitative change in how the junction is perceived. We no longer deal with microns of side-wall and square microns of areal junction, but instead with square microns of core area and square microns of border regions. The border regions will now be lumping side-walls with their neighbouring areal regions which signifies that instead of breaking each pn-junction apart into geometrical units, we will now be considering functional units. The concept is illustrated in figure 6.16 (this can be compared to the equivalent figure for the linear model, found in Figure 6.5 on page 260).

The rest of the processing remains unchanged: We shall use the maximum power output metric and once again define power coefficients, but this time instead of using our previously calculated 'net areal' and 'net side-wall' figures we shall be using 'net core' and 'net border' numbers. These can be derived from the net area and perimeter metrics according to the following equations:

$$C = A - (P - 4dN)d \tag{6.3}$$

$$B = (P - 4dN)d \tag{6.4}$$

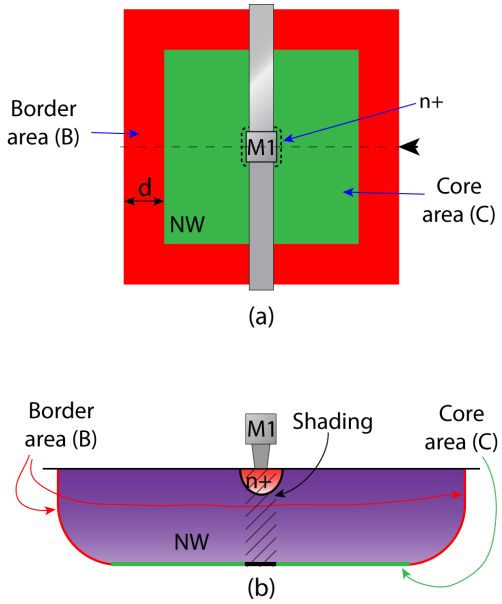

Figure 6.16: Illustration of the concepts of net areal (represented by the green area in (a) and green lines in (b)) and side-wall (shown as red lines) junction components from the perspective of the extended junction decomposition model. (a) Top view of an illustrative, idealised pn-junction with a single metal contact to its N-well terminal and the supply line to that contact. (b) Cross-section of the illustrative pn-junction in (a) along the line implied by the black arrowhead in (a). Noteworthy features: (i) The side-wall junction component now encroaches into what was previously areal junction territory. (ii) The idealised pn-junction sample still has hard corners in its top view. This is because even the extended model does not deal with this issue for the sake of simplicity. (iii) The side-wall component is interrupted by the presence of overhanging metal as seen in (a). (iv) The areal component is interrupted by the presence of overhanging metallisation, but not overhanging diffusion regions (or in general other doped regions) as seen in (b). Abbreviations: NW: N-well. M1: Metal layer 1. n+: n-diffusion.

where C is the net core area, B is the net border area, N is the number of rectangular basic cells (all our basic cells are rectangular[16]).

The determination of the value of parameter d is a matter of consideration. We can begin by observing that it is going to be dependent on the junction type and the manufacturing technology (as these determine the doping profile) and also that it is likely to be of similar or inferior magnitude compared to the minimum design length for each specific combination of junction type and technology.

Setting d also presents us with an interesting trade-off: with larger values of d is, the remaining core area is closer to the ideal, infinite extent areal junction case, but the more device area is relegated into the lumped, border component. Conversely, smaller values of d imply less ideal core areas, but can be used to characterise potentially smaller devices than higher values of d would allow, especially if d is larger than the minimum design length for the junction type and technology under test.

6.5.2 Processing and results with the extended model

In order to explore the possibilities that the extended model offers we have decided to focus our study on the five H18 devices that feature pure N-well on substrate (NS) junctions (no extra dopants in the vicinity). This sets the junction type and technology of operation. The next step is to determine the value of parameter d.

The strategy we have decided to implement for the determination of d relies on computing power coefficients for all pairs of H18, pure N-well on substrate devices on Svejk die #1 (SVJ1-7,8,9,10,11) for different values of d. In this particular case we have a way of finding out whether our choice of d is sufficiently good or not: we can expect that for values of d above a certain limit, all computed coefficients will be roughly the same, regardless of which pair of devices they have arisen from. At that limit value, parameter d has taken its optimal value: sufficiently large to guarantee convergence within tolerance but not unnecessarily large. Should this expectation be indeed confirmed by computed data, then we can state that the variety of coefficients computed in table 6.18 for our H18 test devices were indeed a result of this blur between areal and side-wall regions. If the expectation is instead proven wrong, then the extended model too is insufficient for fully understanding how various component parts of real CMOS junctions contribute to overall performance.

The results of this endeavour are summarised in table 6.25 and visualised in Figure 6.17.

Figure 6.17 shows unequivocally that each value of d leads to different computed power coefficients. Furthermore, not all pairings are affected in the same way. Notably, pairings involving device #7 (SVJ1-7) show a weak trend towards higher core and higher border coefficients whilst all other pairings show a straightforward collapse of the border coefficient as the side-wall is lumped with an increasingly large amount of areal junction.

The trend shown by the larger devices (pairings not including device #7 (SVJ1-7)) is straightforward to explain: At small values of d the side-wall's contribution is added to the contribution of a narrow strip of areal junction and the total power output is divided by the area of the narrow strip. Hence, the

[16] The terms $4b$ in the formulae above generalise for basic cells with arbitrary number of right angle corners to $d(C_i - C_o)$, where C_i is an 'inward' corner (interior of the device lies within the right angle) and C_o an 'outward' corner (interior of device lies within the 270^o angle).

Table 6.25: Maximum power output normalised by junction area or perimeter ('power coefficients') values obtained by pairing pure N-well on substrate (NS) type junctions residing on Svejk (SVJ) dies (Svejk design, dev. #7 (SVJ-7) to Svejk design, dev. #11 (SVJ-11)) in a few, sample combinations under the 'extended model' and for a number of values of parameter d. Standard illumination was used (see table 6.3).

SVJ POWER COEFFICIENT SUMMARY				
DIE	**DEV. PAIR**	**d**	P_{core}	P_{border}
		μm	$pW/\mu m^2$	$pW/\mu m^2$
SVJ1	7,8	0.00	87.86	-0.65
		0.25	88.59	84.28
		0.30	88.86	84.73
		0.40	89.77	85.33
		0.50	92.05	85.81
SVJ1	7,10	0.00	90.61	-1.52
		0.25	92.29	82.24
		0.30	92.93	83.29
		0.40	95.05	84.70
		0.50	100.37	85.81
SVJ1	8,9	0.00	42.03	20.21
		0.25	40.78	132.20
		0.30	40.49	118.72
		0.40	39.86	101.86
		0.50	39.15	91.73
SVJ1	9,10	0.00	41.30	20.93
		0.25	40.00	134.66
		0.30	39.70	120.70
		0.40	39.04	103.24
		0.50	33.93	98.10
SVJ1	9,11	0.00	35.36	26.71
		0.25	34.71	151.31
		0.30	34.57	133.54
		0.40	34.26	111.37
		0.50	33.93	98.10

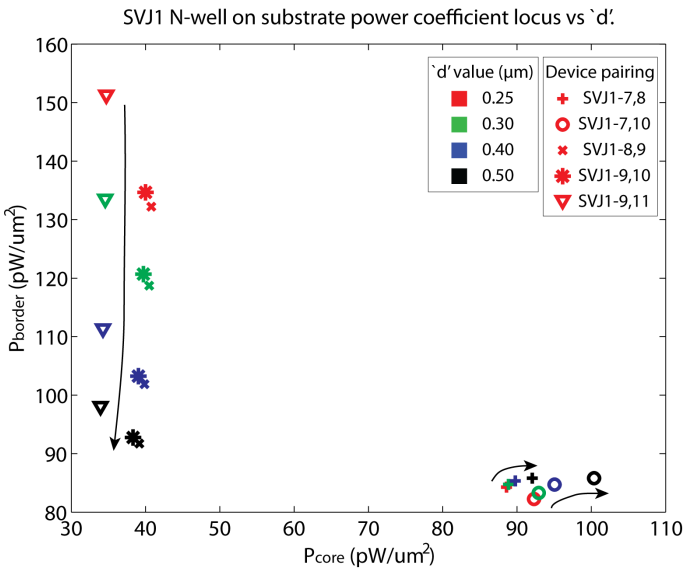

Figure 6.17: Power coefficient locus in the extended model for different values of parameter d. each colour represents a value of d whilst each symbol represents a device pairing from Svejk die #1 (SVJ1). The black arrows show the trends with increasing values of d.

border region overperforms the core by a large margin. As d increases, the contribution of the side-wall is shared with an increasingly large extent of areal junction. As such, the contribution of the side-wall is 'watered down' with increasing valued of d and we can expect that in the limit where $d \to \infty$ both core and border coefficients will settle to one common value. On the other hand, the pairings involving device #7 (SVJ1-7) behave in a manner that is very hard to explain.

More importantly, the predicted convergence of all power coefficients towards a common set of values does not occur. This hints towards the possibility that device #7 (SVJ1-7) behaves in a fundamentally different way compared to its larger counterparts for reasons that we cannot specify.

However, if we decide to relegate results associated with device #7 (SVJ-7) to some random factor that didn't affect all other results, we observe that even with a d value of 0.25 microns the power coefficients for all other pairings show a certain amount of convergence that improves with higher values of d. Thus introducing this parameter does serve to improve the model somewhat, but still leaves a number of issues unresolved, notably the issue of determining an optimal value for d. In the ideal case we already know that the optimal d is found at the minimum value at which all power coefficients from all pairings of devices belonging to the same technology and type converge. In reality there will be variation due to mismatch, measurement noise etc. and indeed we notice that even with the maximum value of d tested, our 'well-behaved' power coefficients didn't successfully converge either. Thus further study is required in order to fully understand and quantify the 'quality' of any value of d for any given junction type and technology and to understand what exactly limits this quality and how.

Finally, we need to mention the constraint imposed upon our choice of d. The value of d is practically lower-bounded by the size of our devices. Notably, device #7 (SVJ1-7), the smallest device used for these computations, is reduced to 100 % border area for $d = 0.5\,\mu m$. This is the reason why the largest value of d appearing in table 6.25 is 0.5 microns. This is yet another reason as to why working with devices consisting of small basic cells is ill-advised for experimental purposes if the behaviour of large

basic cells hasn't been previously elucidated.

6.5.3 Interactions between basic and extended model: case study

In this subection we return to the conundrum of the negative side-wall coefficients computed for some pairings of N-well on substrate (NS) devices on our Svejk family dies (SVJ). The question we will attempt to answer is: if we assume that the pn-junction is described accurately by the extended model, then is there any possibility that the basic model allows negative side-wall coefficients to appear? More specifically, we wish to also know whether the size of the basic cells of the devices involved can affect the sign of the computed coefficients.

In order to tackle the question we begin by setting up the system. We will be working with a pair of devices, #1 and #2 which consist of a single basic cell and are both rectangular. We will then assume that the accurate description of any device x is given by:

$$P_{max,x} = p_{core}C_x + p_{border}B_x \tag{6.5}$$

where p_{core} and p_{border} are the power coefficients of the core and border areas respectively. $P_{max,x}$ corresponds to a measured value and is assumed to be indisputably correct for this set of calculations. Meanwhile, our basic model shall attempt to describe the situation according to:

$$P_{max,x} = p_{areal}A_x + p_{fringe}P_x \tag{6.6}$$

If we take devices #1 and #2, and apply both the assumed, extended model and the basic model we obtain the following system:

$$P_{max,1} = p_{core}C_1 + p_{border}B_1 = p_{areal}A_1 + p_{fringe}P_1 \tag{6.7}$$

$$P_{max,2} = p_{core}C_2 + p_{border}B_2 = p_{areal}A_2 + p_{fringe}P_2 \tag{6.8}$$

We can now solve for the side-wall power coefficient p_{fringe} in order to obtain:

$$p_{fringe} = \frac{A_1p_{core}C_2 + A_1p_{border}B_2 - A_2p_{core}C_1 - A_2p_{border}B_1}{A_1P_2 - A_2P_1} \tag{6.9}$$

However, the area, perimeter, core and border of any device are interlinked via equations 6.3 and 6.4. Using these we can express p_{fringe} as a function of only areal junction extent and perimeter (A and P), i.e. eliminate core and border area extents from the system:

$$p_{fringe} = \frac{d(p_{border} - p_{core})[A_1(P_2 - 4d) - A_2(P_1 - 4d)]}{A_1P_2 - A_2P_1} \tag{6.10}$$

where parameter d has now appeared explicitly in our calculations.

We notice that a constraint of the system is that both perimeters of our devices need to be larger than $4d$, otherwise the devices are reduced to featuring below-zero core areas. Additionally, the expression is symmetric with respect to A_1, A_2 and P_1, P_2 so that exchanging the indices on A_x and P_x results in the same expression. This simply confirms that the order in which we choose the devices to be compared is irrelevant; it only matters which devices we choose. Finally, quite notably the relation between the border and core area power coefficients also determines the sign of the fringe coefficient in the basic model although the rest of the expression doesn't clarify in which direction.

The right hand side of equation 6.10 does allow conceivably both negative and positive side-wall coefficients to appear for the same junction type in the same technology solely as a result of the geometry of the devices involved. Therefore we will examine two interesting cases that correspond to the Svejk family devices that were studied in table 6.18 on page 263, namely Svejk design ,devices #7 - #11 (SVJ-7,8,9,10,11).

In the first case we shall consider a pair of devices described thus:

- Device 1 is a square with side a.
- Device 2 is a rectangular device with sides a and $b = 2a$.

These devices will correspond to Svejk design, devices #7 and #8/10 (SVJ-7 and SVJ-8/10) respectively. Replacing the areas and perimeters of the devices with the values derived from their geometrical data we obtain:

$$p_{fringe} = d(p_{border} - p_{core})\frac{a - 2d}{a} \tag{6.11}$$

whose sign is determined by the relation between the core and border power coefficients since $\frac{a-2d}{a}$ is necessarily positive or device 1 would be reduced to featuring a negative core area.

In the second case we shall consider a pair of devices described thus:

- Device 1 is a rectangular device with sides a and $b = 2a$.
- Device 2 is a square with side $2a$.

These devices will correspond to Svejk design, devices #8/10 and #9 (SVJ-8/10 and SVJ-9) respectively. Replacing the areas and perimeters of the devices with the values derived from their geometrical data we obtain:

$$p_{fringe} = d(p_{border} - p_{core})\frac{a - d}{a} \tag{6.12}$$

whose sign is again determined by the relation between the core and border power coefficients as $\frac{a-d}{a}$ is necessarily positive.

The implication is that the computed values of p_{fringe} will in both cases have the same sign. Thus we have proven that if the underlying behaviour of our N-well on substrate (NS) pn-junctions in technology H18 is assumed to be accurately described by our extended model, then the fact that we obtain negative side-wall power coefficients for the Svejk design, device #7 and #8/10 pairings (SVJ-7,8/10) and positive values for the device #8/10 and #9 pairings (SVJ-8/10,9) is not explainable by considering the discrepancies between the basic and extended models alone.

Note: In both these cases a border power coefficient that is higher than its core area counterpart tends to render the side-wall power coefficient positive.

6.5.4 The extended model: summary

The implications stemming from our findings are severe: A simple re-configuration of our basic 'areal + side-wall' model into a more advanced 'core + border' model, though an improvement over the basic model, is insufficient for explaining the full complexity inherent in realistic pn-junctions in CMOS. Furthermore, even if a suitable value of d had been found in the computations conducted for this section it would not generalise well across technologies and junction types.

Therefore, it is unfortunate that further experiments need to be conducted with larger sets of test devices (pairs are simply insufficient for this task), ideally using larger basic cells, before we can determine suitable values of parameter d for each junction type and each technology under study.

6.6 Additional considerations

Throughout this chapter we have studied the behaviour of our designed pn-junction devices in the capacity of optical power harvesters. Though every care has been taken to ensure that the measurements involved were as accurate as possible there are always uncertainty factors that cannot be easily eliminated. Some issues such as the blurring that occurs between side-wall and areal junction components have already been discussed (6.4), but others have not. In this section we mention these extra uncertainty factors and attempt to understand their effects on the experimental set-up.

Linear model validity

Throughout our discussion we have implicitly assumed that the maximum power delivery for each junction can be linearly decomposed into contributions from each component, areal and fringe. I.e.:

$$P_{max} = P_{areal} + P_{fringe} = Ax + By \tag{6.13}$$

where x and y are the design area and perimeter values for the device under consideration whilst A and B are the normalised power output coefficients under standard illumination for each square micron of areal junction and each micron of side-wall perimeter respectively.

This is not strictly true, however. In order to illustrate the imperfection we shall consider simplified case where areal and side-wall junction components are geometrically separable. In this case we can model

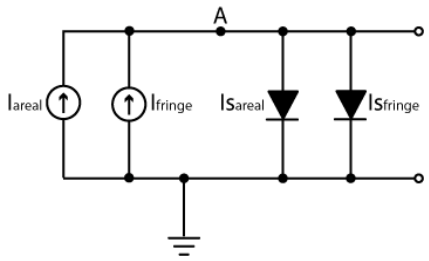

Figure 6.18: Simplified 'areal plus side-wall' photodiode macromodel. Photocurrent I_x always belongs to the device with reverse saturation current Is_x, where x is the photodiode identifier.

our CMOS device as a pair of parallel-connected photodiodes with different reverse saturation currents (see Figure 6.18). Each photodiode will also feature different photocurrent generation capabilities under given illumination conditions (exposed to the same irradiance).

Kirchhoff's current law at node A in Figure 6.18 for some unknown, external load connected between node A and GND takes the form:

$$I_{load} = I_{areal} + I_{fringe} - (Is_{areal} + Is_{fringe}) \left(e^{V/nV_T} - 1\right) \tag{6.14}$$

where I_{load} is the load current, I_x is the photocurrent for device component x ($x \neq$ '*load*'), Is_x its reverse saturation current, V the potential difference between node A and GND, n the diode ideality factor and V_T is the thermal voltage.

We know that the reverse saturation current depends on the semiconductor properties of the diode and is proportional to its area. In the case of side-wall junctions the 'junction component width' is fixed, so the reverse saturation current will be proportional to the perimeter of the CMOS junction to which they belong. Reasoning similarly we can find that photocurrent will be proportional to the area of the areal component and the perimeter of the side-wall component. This allows us to express photocurrent under given illumination conditions as:

$$I_{areal} = k_{areal} Is_{areal} \tag{6.15}$$

$$I_{fringe} = k_{fringe} Is_{fringe} \tag{6.16}$$

where k_x are proportionality constants.

By introducing 6.15 and 6.16 in 6.14 and rearranging we obtain:

$$I_{load} = (k_{areal} + 1)Is_{areal} + (k_{fringe} + 1)Is_{fringe} - (Is_{areal} + Is_{fringe})e^{V/nV_T} \tag{6.17}$$

i.e. the sum of the photocurrents minus the sum of the currents draining down the forward biased diodes. Notably, load current is a linear combination (specifically the simple sum) of the contributions from each junction component (component photocurrent minus component drain across the forward-biased diode).

The expression for power delivery to the load then becomes:

$$P_{load} = V\left((k_1+1)Is_1 + (k_2+1)Is_2 - (Is_1+Is_2)e^{V/nV_T}\right) \tag{6.18}$$

where the subscripts have been shortened from 'areal' to '1' and from 'fringe' to '2' for clarity.

We then compute the voltage bias at which power is maximised by solving the equation:

$$\frac{\partial P_{load}}{\partial V} = 0 \tag{6.19}$$

which after rearranging yields the following implicit relation:

$$\left(1 + \frac{V}{nV_T}\right)e^{V/nV_T} = \left(\frac{(k_1+1)Is_1 + (k_2+1)Is_2}{Is_1+Is_2}\right) \tag{6.20}$$

Equation 6.20 shows that the voltage at the maximum power point depends on the areal/side-wall make-up of the junction as express by their reverse saturation currents for any $k_{areal} \neq k_{fringe}$.

This result is important because we have already seen that the load current is a linear combination of areal and side-wall contributions. Therefore, maximum power output, which is the total load current multiplied by the voltage at the maximum power delivery point is not a linear combination between junction component contributions. In our analysis we have ignored this effect because of the relatively similar voltage outputs obtained from junctions of difference areal/side-wall preponderance but the same type.

As an example of this phenomenon we note open circuit voltages for Svejk design, devices #7 to #11 (SVJ-7 to SVJ-11) from both Svejk die #1 and die #2 in tables 6.5 and 6.6. By setting the load current at $0\,A$ in 6.17 and solving for V, it can be proven that the open circuit voltage can be expressed in our simplified model as:

$$V_{OC} = nV_T ln\left(\frac{(k_1+1)Is_1 + (k_2+1)Is_2}{Is_1+Is_2}\right) \tag{6.21}$$

where V_{OC} is the open-circuit voltage and is again independent of the make-up of the pn-junction only for different values of k_{areal} and k_{fringe}.

Notably, open-circuit voltages for devices with identifiers #8 and #10, which feature the exact same layout geometry and only differ in the orientation of the pattern (i.e. which way the rectangular basic cells face), seem to be much more similar than those of devices with different basic cell structures, e.g. a #7 design with a #11 design. However, the differences are rather small. Open circuit voltages differ by less than 2% in the Svejk designs #7 to #11 (SVJ-7 to SVJ-11) whilst the ratio of areal to side-wall junction components ranges between 0.316 microns for Svejk design #7 (SVJ-7) to 2.567 for Svejk design #11 (SVJ-11).

Note: Even though we used the full model for diode current we could have easily used the simplified, $I_{diode} = Ise^{V/nV_T}$ model. The results would show little change as we can assume that $k_x \gg 1$ for any useful level of illumination.

Finally, we must for the sake of completeness address the fact that we implicitly assumed that the photocurrent of each junction component is independent of the other component. This doesn't strictly

hold true because photocurrent is a function of depletion region size (ideally direct proportionality), which itself is a function of voltage across the junction (rough square root dependence). However, modulation of the depletion region width is negligible when voltage variations are so small and additionally, any attempts to take the phenomenon into account might need to take into consideration the fact that depletion regions are not clear-cut regions of semiconductor volume, but rather are characterised by a free carrier concentration gradient (in the case of ideal, abrupt, homogeneous pn-junctions exponential free carrier concentration as a function of distance between the edges of the depletion region - itself only an approximation). For more information reading the relevant sections from [1] is recommended.

Metallisation shadow casting

Every semiconductor device must be connected to other devices or pads via metal tracks and metal-semiconductor contacts. These tracks will cast shadows upon the chip front-end. The same occurs with intentionally designed metallisation sheets that act as light-stopper masks. The effects of the shadows that these structures cast have already been considered by using net junction area and perimeter figures, but this implicitly assumes that the incoming light falls upon the die perfectly perpendicularly.

Every care was taken to ensure that each PCB lies flat on top of the micro-positioner platform and that each packaged die is firmly inserted into its corresponding package holder. Furthermore, the light source was aligned to the best extent so that it points perpendicular to the surface of the micro-positioner platform, but small angle deviations may still be present. Finding out the extent of these positioning errors is not trivial.

If a collimated beam falls perpendicularly upon an object of width d, then it will leave a shadow of width d on a screen positioned behind the object and perpendicular to the direction of the beam. The same will hold if the beam falls upon the object at any angle up to and not including 90^o (we define 0^o the point where the beam travels perpendicular to the screen) if the object casting the shadow has zero thickness. When it comes to metal tracks that is not the case with tracks designed at minimum allowable width featuring thicknesses that are comparable to their width. If we assume a situation where the metal tracks are $0.25\,\mu m$ wide and $0.25\,mum$ thick, then at 45^o light angle 'error' the shadow cast upon the underlying screen will be approx. $0.35\,\mu m$ long, or about 40% extra, which is the worst-case scenario. Thus, higher angles tend to increase the amount of shading falling upon devices. So long as overall coverage of the device by metallisation is kept low and the angle errors also reasonably low (few degrees at most), this should not pose a significant problem.

In our set-up we do have metal tracks that are close to or at minimum design width but we have no information on the angle at which light beams enter the system. Nevertheless we believe that the effect that causes shadows to cover more of our designed devices will not play a very significant role due to the small coverage of each device by metallisation (data not shown).

More importantly, beam entry angle errors will cause the shadows to shift their target location on the screen by $l \cdot \tan(\theta_{err})$ where θ_{err} is the error angle and l is the distance of the object from the screen. Because different metal layers lie at different distances from the surface of the die their targets will shift by different degree for the same angle error. Typically metal layers will be suspended up to a few microns above the surface of the die. Further complications arise when considering that pn-junctions are not just superficial front-end features, but rather have a certain depth. This degree of complexity means that a

vast amount of work has to be performed in order to determine the effects of beam angle errors on the illumination of the entire pn-junction volumes. In our set-up we have no way of knowing exactly since the required analysis was not performed.

Measurement noise

We have already alluded to the effects of noise on measuring photocurrents and even seen an example of measurement noise on swept I-V curves (6.4). The obvious way of eliminating such effects would be to take many I-V sweeps of each device and then average the obtained values. This was not done, but in retrospect should have been done. For that reason in this section we shall discuss what we expect the benefits and pitfalls of such approach to have been, had it been followed.

When taking multiple measurements one would expect noise contributions of high frequency to disappear and only uncertainties that operate on longer time scales to remain. Thus the highly oscillatory results obtained by generating single I-V sweeps one point at a time could be averaged out. Nevertheless, changes in ambient temperature, the quality of illumination, the device alignment and positioning might still leave a mark. We would expect these effects to have been particularly pronounced in the case of taking repeat measurements at long time intervals.

Therefore the best practice would have been to modify the automated MATLAB scripts used to extract I-V sweep data from each device so that measurements are taken from each device multiple times within a small interval of time. At the end a 'dump' containing the raw measurement is saved should result and a final, noise-cleansed I-V trace can be displayed.

At this point we have two practical implementation options: a) Take a succession of I-V sweeps and average them. b) Measure photocurrent at each bias point repeatedly before moving on to the next bias point. Option a) effectively generates the average I-V sweep by superposing homologous data-points from each individual sweep that have all been taken at intervals of time long enough to allow the processing of one entire I-V sweep. option b), on the other hand is based on averaging data-points taken in very quick succession, with minimum delay between them. Thus, in option a) we obtain an averaged I-V sweep where each data-point of the averaged I-V potentially 'contains' more spread (i.e. has been extracted from data with higher spread) but also where the 'average ambient conditions' (AAC) experienced by each point of this averaged I-V sweep are similar; that is to say if we by some measure averaged the ambient conditions present during the measurement of each data point that composes an averaged I-V data point and compared them to the obtained AACs for each data point of the averaged I-V sweep we would notice that they are fairly similar. On the other hand, option b) yields an I-V sweep where each data-point of the averaged I-V was obtained under much more consistent conditions and therefore exhibits lower spread. The price for this is that the averaged I-V data points obtained towards the end of this procedure may have been obtained over substantially different AACs compared to those obtained towards the start.

Finding how methodologies a) and b) differ in the quality of the results they yield would be a good subject of further study. Chances are that we would have chosen to implement method a) because then each point in the obtained averaged I-V would have been measured under very similar average conditions.

Other effects

Previously we have considered the effects that slow-acting phenomena can have on repeated measurements of the same device. Such effects, with ambient temperature variation figuring as a prime example, will also corrupt results taken from different devices. Moreover, unavoidable alignment errors penetrating the system (the system has to be realigned between measurements, unlike when measuring the same device repeatedly) will further degrade results. Finally, effects that we can call 'spatial noise', such as process variation and chip warping will make a further additions to the collection of sources of uncertainty.

These effects can be countered to a certain extent. For example, we chose to take measurements in very quick succession with all Ninja (NIN) and Teddy (TED) power recovery measurements being taken within one day and all Svejk (SVJ) measurements within another. Alignment errors were at least partially resolved by always searching for other, higher photocurrent maxima once what initially appeared to be the global maximum had been reached. This confirmatory action most often did not lead to alignment changes but was deemed a necessary precaution. Finally the issue of 'spatial noise' cannot be solved without resorting to using more dies (in order to gain an understanding of the 'average' Silicon conditions prevailing at the location where each device is placed) and fabricating devices whereby the same devices change positions on the die (in order to assess exactly how the specific location on the die affects device performance).

Note: Certain factors, such as scatter within the back-end of the die have been assumed to be insignificant. Lack of reliable data and time to perform the relevant analysis renders proving this claim difficult, but the fact that CMOS imagers generally don't seem to be limited by scatter in the back-end of the process is a good indication that we can discount the effect.

6.7 Power management circuit design

A power management unit (PMU) has been designed as part of this project. While the design of the PMU was not innovative in itself, the aim was to see whether a simple, rugged design of a PMU could be ran at a sufficiently low power budget so that the power harvesting structures could theoretically service it and at the same time generate a sufficient excess of power output that could be used to drive other circuits on the die (e.g. a sensor node).

Note: Schematics are given here in brevity with only the important aspects in their design highlighted. Full diagrams including devices sizes are to be found in the appendix (chapter E).

6.7.1 Design overview

The entire system features only one input and one output. The input port links to a photosensitive element that acts as a power harvester while the output provides a stable, regulated voltage that can act as a power source for any load circuit the die may be hosting. Naturally, a ground connection also exists. For simulation purposes the input node will be connected either to an ideal voltage source (preliminary experiments), or to a standard photodiode macromodel where the current source is set to a constant level in order to simulate conditions of constant illumination (advanced tests). The shunting

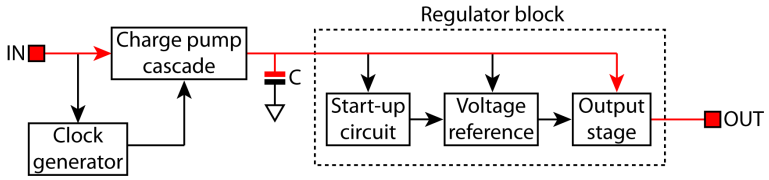

Figure 6.19: Block diagram of the designed PMU. The capacitor simply marked as 'C' is the smoothing capacitor between the charge pump output and the regulator block. The main power path has been highlighted in red as opposed to other paths that feed components of the PMU and are thus encroaching on the DC-DC conversion efficiency factor. Input and output terminals are marked with red squares.

resistor will be ignored and the series resistor considered negligible (good quality metallic connection between photodiode and system input devices assumed). At the output, a simple current source load will simulate a circuit that exhibits constant energy demand. Of course more detailed and inclusive simulations would be necessary to determine the behaviour of the system within a real set-up, but such detailed investigation is not within the mandate of this project.

Internally, the system is subdivided into a number of standard functional modules that will be considered individually throughout this section. Conceptually, the system operates on the basis of a low-voltage clock generator, fed directly by the unregulated voltage at the system input. The clock drives a multi-stage Pelliconi cell charge pump [6, 7] that upconverts the voltage to a fluctuating value lying comfortably above the target regulated value. These fluctuations are impossible to eliminate completely, but a smoothing capacitor at the end of the charge pump cascade does alleviate the issue somewhat. This fluctuating, unregulated voltage then feeds into the regulator block that itself consists of three sub-units: a start-up circuit, a voltage reference and an output stage that attempts to maintain the regulated output voltage at a fixed value. All this is visible in the block diagram provided (Figure 6.19).

In terms of design specifications, the system should receive a 'realistic' $0.5\,V$ stable power supply and provide at least $12\,\mu A$ at a $1.5 - 1.9\,V$ stable voltage level. The choice of load current capability, though largely arbitrary, was based on the multitude of IC designs that can perform simple functions with micro-power and below. A stable voltage supply was defined as featuring a ripple of less than $10\,mV$. These are the only specifications imposed upon the design along with the implicit requirement of maximising efficiency (or conversely minimising power dissipation within the PMU).

6.7.2 Clock generator

The generator module is a very simple design consisting of a 21-stage ring oscillator and a vast array of fanout (i.e. 'drive strength boost') inverters aimed at creating a strong enough clocking signal to drive the charge pump.

The ring oscillator was designed to contain 21 inverters, a number chosen because it was empirically found out that at that number of stages the frequency generated at the target operating input voltage of around $0.5\,V$ it yielded a clock frequency that was neither too low to allow the charge pump to shuttle enough charge to the regulator block, but neither so high that the system would have to draw unrealistically high amounts of current while maintaining the target input voltage. Note: the charge pump average current intake (and consequently also its output drive strength) depends linearly on the frequency at which it is clocked. For that reason balancing PMU current intake demand so that it

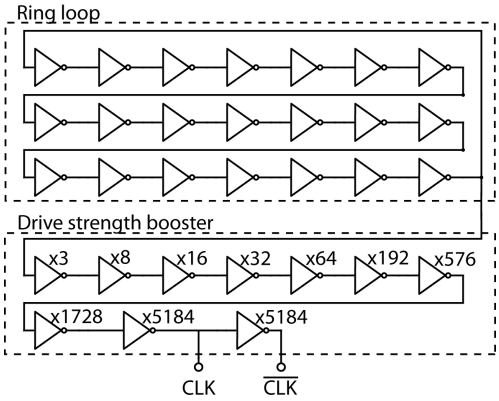

Figure 6.20: Clock generator module including the ring oscillator loop and the drive strength booster cascade. The power supply connection is not explicitly shown, but power is drawn directly from the output of the power harvesting element.

matches the power harvester's capability to provide and leaves enough available excess power at the output is no trivial task. In the literature review section there has been a discussion about specialised systems that attempt to find that balance dynamically via relatively simple hardware-only processing (the so-called maximum power point tracking systems (MPPT systems) -see section 2.4.2-).

Naturally, just like with any other digital circuit, the clock generator system needs to be able to drive its target circuit, which in our case is the charge pump. The charge pump receives its clocking directly to an array of relatively large capacitors (by IC design standards) that are tasked with upconverting large amounts of charge and shuttling them to the system exit. Thus, the loading at the output of the clock generation module is by no means small. The ring oscillator itself, a subsystem consisting of minimum-size, balanced devices can not drive such load at its own output frequency. For that reason an array of drive strength boosting inverters was implemented.

Once again, largely empirical data and the well-known rule that the optimal drive strength amplification per stage factor takes its optimal value in the region around 2-4[17] we decided to implement a drive strength booster containing nine stages for a total current drive amplification factor of approximately 5184. The stage amplification factors were as follows: 3, 8/3, 2, 2, 2, 3, 3, 3, 3. Original plans were to implement the entire cascade with $\times 2$ stages, but layout and power dissipation considerations forced more economical use of space. This occurred at the cost of sharp transitions at clock signal edges.

Finally, at the end of the booster cascade another inverter block with the same drive strength as the final stage was added in order to create the complementary phase clock signal (CLK bar). The clock (CLK) and clock bar signals were therefore non-overlapping, however no such signals were requested in order to run the charge pump[18]. Note: the inverters generating CLK bar can drive their targets more strongly than their counterparts tasked with the generation of CLK because the CLK generators need to drive the inverters that give rise to CLK bar in addition to their actual targets within the charge pump. These charge pump targets represent the same load for both CLK and CLK bar generator. Schematics for the clock generator module can be seen in Figure 6.20.

[17] See [8] for more details to this oversimplified rule of thumb.
[18] Unlike in series-parallel charge pump designs where non-overlapping clocks are crucial.

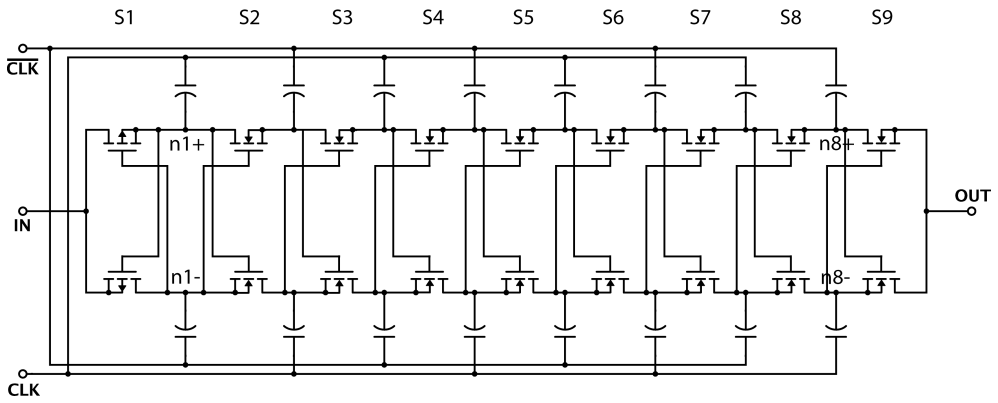

Figure 6.21: Charge pump design used in the PMU. The SX indicators, where X is an integer between 1 and 9 indicate the transistor pair stage number. Note the different connectivity pattern for the gates of transistor pair S1. 'nY+' and 'nY-', where Y is an integer between 1 and 8 are node names.

6.7.3 Charge pump

Our choice of charge pump design, the Pelliconi cascade, was made on the basis of its functional, symmetric design and simple biasing scheme. The symmetry of the design arises from the fact that the charge pump assembly consists of two sub-units that operate in a time-interleaved formation. What this means is that while one sub-unit is in its charge-up phase (preparing to voltage-upconvert an amount of charge), the other is in its discharge phase (releasing already up-converted charge). The net result is that the common output of each sub-unit is constantly being serviced by one of the two sub-units. Of course, other designs either include or can be adapted to include this capability. The Pelliconi cascade, however, features the additional advantage that it does not require non-overlapping clocks in order to operate. Unlike series-parallel configurations where overlapping clocks lead to dissipation of the accumulated charge, the Pelliconi cell employs diode-connected transistors that act as rectifiers and prevent excessive 'backwards leakage'[19] of charge during those times when clock overlap forces all stages in the cascade into the same state (charge or discharge state). In the ideal case, alternate stages are always in different states.

A schematic of the charge pump design used is shown in Figure 6.21. Note: the S1 transistor pair consists of nMOS devices while all other pairs consist of pMOS devices. If pair S1 consisted of pMOS devices, they would have had to be diode-connected in order to keep the pump operational. The implication of this design modification would be that the S1 pair's gates would never be able to exploit the above-VDD voltages that appear at the nodes marked as 'n1+' and 'n1-' in Figure 6.21. In the current configuration when an S1 device needs to switch on, it benefits from an anove-VDD voltage at its gate and thus reduces voltage loss across the stage due to the threshold voltage effect.

When designing a charge pump an important decision concerns the number of stages to be used. The trade-off is between raising the output voltage enough above the target regulated voltage in order to ensure proper regulator block operation and raising it sufficiently little in order to avoid excessive losses. These losses do not only encompass dissipation arising from down-regulation, but also include the power consumed to run potentially unnecessary stages within the charge pump itself (more stages imply more capacitance to drive overall).

[19]Charge leaking 'backwards' towards the beginning of the cascade. A phenomenon that should ideally not occur.

The issue of choosing the correct number of stages, however is further complicated by the fact that the average load current affects the output voltage of the charge pumping system. It is outside the mandate of this thesis to explain the exact relation between output voltage and load current, but a brief explanation is given for convenience. There is a theoretical limit to which the output voltage of a charge pump can be pushed (we will call this the 'no load limit voltage' of the charge pump). This is achieved for a purely capacitive load. Over time, the load capacitor accumulates charge and thus the output node voltage vs GND rises. In the region of the no load limit, the voltage difference between the output node and the nodes before the final stage (in Figure 6.21 labeled as 'n8+' and 'n8-') is so small that no more charge can enter the load capacitor. At that point the charge pump simply toggles between states without shuttling any net charge in any direction. Note: as the output voltage reaches its theoretical upper limit the charge that can be added to the load capacitor in each cycle diminishes. This occurs because charge accumulation at the output node is implemented by using charge sharing between the output node and the nodes between the final stage. The result of this is that the voltage tends towards its limit asymptotically.

If a constant load current is added to the output node, then some of the accumulated charge 'leaks' away every half-cycle. Provided that this current is weak enough, the charge leaked over any half-cycle 't' can be at least replaced by the influx of new charge from the capacitors preceding the final stage transistors. As a result, the final, average output voltage value will drop from its theoretical upper limit down to the point where the charge injection capability per half-cycle of the pump balances the average current drain at the output node. The average output voltage over a cycle, however, will still tend asymptotically to the new equilibrium value. Thus, it is important to note that any obtained results used to characterise nominal charge pump behaviour must be taken at 'steady state', i.e. when the system has reached equilibrium. The asymptotic nature of convergence to equilibrium, nevertheless, means that small errors will always persist.

On the other hand, there is a critical output voltage value below which the final stage capacitors fully discharge onto the load capacitor. Determining that value, just like determining the theoretical upper limit, is not trivial, especially if transistors are modeled as full analogue devices rather than merely switches with a threshold[20]. The key issue is that if the output voltage drops below that critical value the capability of the charge pump to inject charge into the load capacitor no longer increases.

One more important aspect that determines the current drive capability of the Pelliconi charge pump is the size of the capacitors used throughout the cascade. Whilst the observations above link current drive capability to output voltages, the size of the capacitors involved are effectively determining absolute current drive capability. Intuitively, a charge pump with capacitors of size 2C should be equivalent to two charge pumps with capacitors of size C working in parallel.

Given this information on the function of the Pelliconi charge pump we can infer that if we know the peak, cycle-average current drain at the output node we can use that value as a constant current draining source at the output of our system for worst-case scenario testing. In this project we chose a load current of $25\,\mu A$ as this constant current drain value. Therefore, given the system requirement of at least $12\,\mu A$ current drive capability we have implicitly budgeted a current consumption of $13\,\mu A$ for the regulator block (voltage reference and output stage). Empirically it was then found that nine stages

[20] The critical value will depend on the ratio of the values of the load and final stage charge pump capacitances and the characteristics of the transistors linking them.

with $\approx 46\,pF$ capacitors was a reasonable design for the charge pump.

The choice of transistor sizes concerns driving the RC constant at each stage node (nX+ and nX-) towards the technology limit given by a single transistor while trying to minimise leakage current. Further details lie outside the scope of this thesis.

6.7.4 Smoothing capacitor

A small smoothing capacitor was added at the output of the charge pump. The capacitor measured $810\,fF$ and should dominate the load capacitance. Given the discussion in the charge pump design section (6.7.3) the load capacitance plays a very important role in the function of the system. Its size will determine how charge injection and ejection affect node voltage. Very large load capacitances will make the system sluggish in terms of start-up time (initial charge-up) and consume valuable layout area. Very small load capacitance can act as a bottleneck that limits how much charge can be transferred from the charge pump final stage capacitors to the load. The details are outside the scope of this thesis, but the net effect is that as the load capacitance decreases, the drop in steady-state charge pump output voltage caused by a constant current drain becomes more extreme. Similarly, the critical voltage value below which the charge pump cannot provide more current decreases. If this is taken to the extreme, then it may well become the case that for a large range of load currents the charge pump is more than able to replenish the charge lost in each cycle, but the resulting steady-state output voltage remains below the usable range, thus wasting a lot of the charge-shuttling capacity of the pump.

Originally the importance of the smoothing capacitor was underestimated and the value was kept low in order to speed up simulations. In hindsight, the load capacitor used for our PMU should have been significantly larger.

6.7.5 Start-up and voltage reference circuits

It is very convenient to study the very simple start-up circuit in tandem with the voltage reference as the functions of the start-up and voltage reference circuits are highly interconnected.

The start-up circuit is a fairly standard design that can be seen in [8]. From the schematics in Figure 6.22 we can see that its principle of operation hinges on shorting nodes ‘a’ and ‘b’. This implies that once the voltage at ‘VDD unregulated’ reaches at most the sum of one pMOS plus one nMOS threshold voltages, then either the top current mirror (gates connected at node ‘a’) or the bottom one (gates connected at node ‘b’) will engage. This ensures that as the voltage on ‘VDD unregulated’ keeps increasing to the minimum operational value the reference will keep conducting current down its main current mirror branches. Once the bottom current mirror fully engages, the start-up circuit disengages and nodes ‘a’ and ‘b’ become electrically separated once again.

The voltage reference circuit is a simple, rugged beta multiplier with a pair of diode-connected transistors acting as the load and a smoothing capacitor. The principle of operation revolves around generating a reference current and then mirroring it out to a consistent load. [8] contains the schematics and all basic details of its design methodology and operation, but a brief overview is given here for convenience. The bottom current mirror pair consists of devices of differing effective widths (drive strengths). Therefore when the same, non-zero drain current passes through them their V_{GS} values are

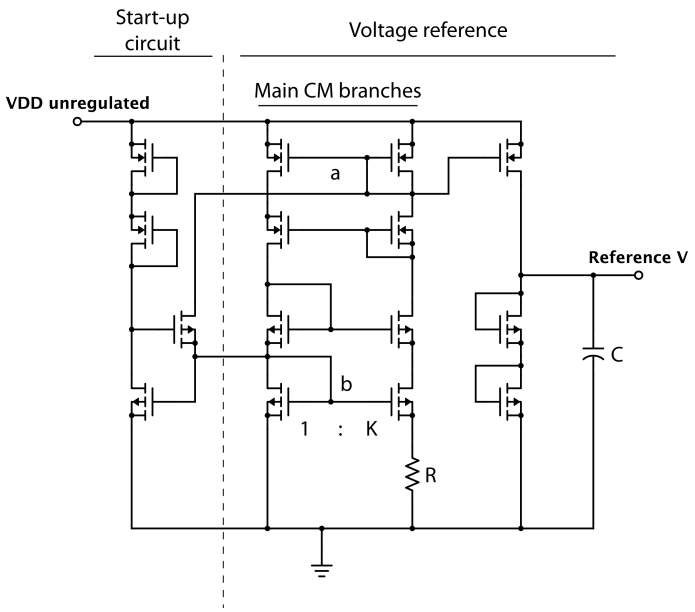

Figure 6.22: Voltage reference and associated start-up circuit. The system 'input' consists of the unregulated power supply whilst the output should be a stable voltage. CM: current mirror.

different (discounting channel length modulation effects). Source degeneration at the wider member of the pair means that there are at least two equilibrium operating points at which the same current can be forced down the drains of both devices. At one equilibrium point (trivial case) the drain currents are both zero. At the other equilibrium point the difference in V_{GS} prevailing because the same drain currents are being forced down two devices with different widths is perfectly compensated for by the source degeneration resistor.

In our design the effective width ratio between the beta multiplier transistor pair that generates the ΔV_{gs} (K in Figure 6.22 was chosen to be equal to 2. The resistor value was $\approx 7.65\,k\Omega$ whilst the smoothing capacitor was $\approx 15.33\,pF$. These values were derived through a mixture of hand calculations and empirical testing deemed adequate for a sufficiently well performing voltage reference. A schematic is shown in Figure 6.22.

6.7.6 Regulator output stage

The regulator output stage consists of a slightly modified Op-Amp with Miller and nulling resistor compensation, driving an output MOSFET. The key metrics that describe the Op-Amp are its gain-bandwidth (GBW) product, phase margin, PSRR and the total current consumption. We shall examine the values for the designed system as a whole in the simulated result section (section 6.8). This means that the Op-Amp plus output nMOS will be considered as part of one module and the source of the output nMOS (the regulated output node) is the node where metrics such as GBW will be determined. This method is chosen in the interests of brevity.

The GBW value was determined mainly by the constraint of using as little power as possible. As such, it was proven impossible to extend GBW into the MHz range without using tens of microamps bias current. A compromise had to be made. This affects the ability of the regulator output stage

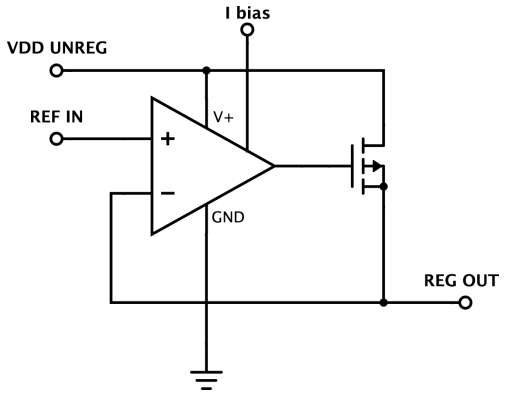

Figure 6.23: Output stage of the regulator block using standard, Op-Amp-based design with a wide output transistor.

to counteract fluctuations appearing at the input arriving from the voltage reference as well as sudden voltage fluctuations appearing at the regulator output node (because of dynamic changes in the effective load at that point).

The aforementioned modifications to the Op-Amp simply included a smoothing capacitor across its power supplies ($\approx 4.16\,pF$), another smoothing capacitor at the Op-Amp output terminal (again $\approx 4.16\,pF$) and a simple diode-connected transistor linking directly to a pad in order to enable the generation of the Op-Amp bias current directly from an external current source. Quoted performance values are given for the system including these modifications. The high-level schematic of the output stage can be seen in Figure 6.23.

The output transistor had an aspect ratio of $\frac{1\,mm}{0.7\,\mu m}$ in order to ensure that the output voltage of the Op-Amp will typically not need to climb too far above the regulated output target voltage (preventing a potential output voltage swing bottleneck issue).

6.8 PMU in operation: simulated results

Throughout its design phase the power management unit had to be subjected to various simulations both in terms of individual parts and in more complex modules and as a whole system. In this section we display the results of the full system simulation, the less crucial module-level simulations having been relegated in chapter C of the appendix. Finer details of the testing/simulation procedure are also relegated in the same chapter of the appendix.

6.8.1 Full system simulation

The entire PMU system can be simulated with VDD, regulator output stage bias current (I_{bias}) and loading conditions. For the purposes of testing we used nominal VDD of $0.5\,V$, I_{bias} of $600\,nA$; a value set to generate the nominal $8.2\,\mu A$ output stage bias current described in the relevant section, and we used as load a current source drawing constant current at $12\,\mu A$. This test configuration we will call the 'standard ideal' configuration in an allusion to the ideal voltage source used to simulate the power scavenger element. The results can be seen in Figure 6.24.

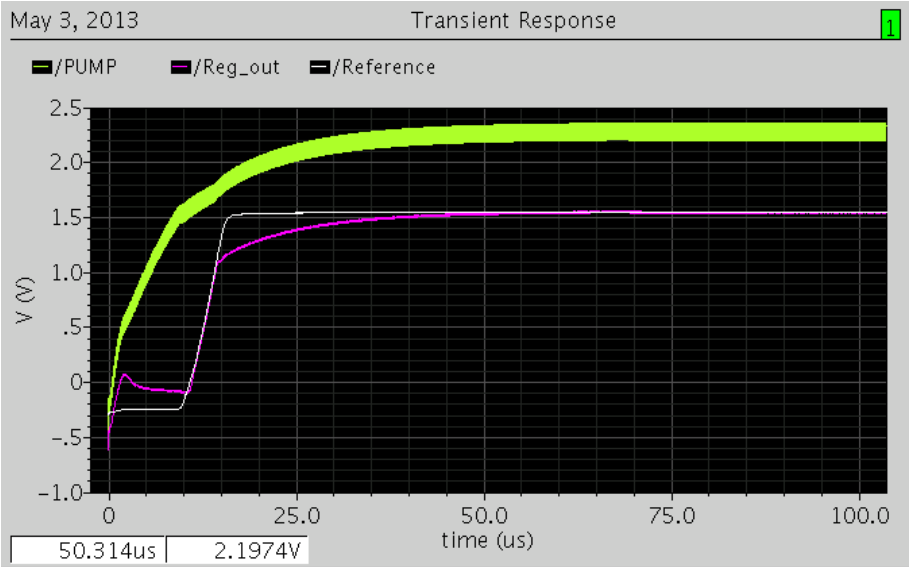

Figure 6.24: Overall PMU behaviour from start-up to steady state. Superimposed traces of voltage as a function of time at the following nodes: Green: Output of the charge pump. White: Voltage reference output. Pink: Regulated output. The test was performed under $0.5\,V$ VDD with a constant current load drawing charge out of the regulated output node at the rate of $12\,\mu A$.

What this result shows us is that the system can hold a regulated voltage level at approx. $1.5\,V$ even when we draw out of it a load of $12\,\mu A$ earmarked specifically for load circuits and at the same time we feed the voltage reference and the Op-Amp that controls the output stage of the regulator block with their respective energy requirements. In other words, the limited current drive capability of the charge pump is sufficient to provide enough current for both the regulator block operation and feed a load of $12\,\mu A$. Caveat: the $600\,nA$ current used as a reference for the generation of the bias currents used in the Op-Amp at the output stage of the regulator block are not included in this calculation. The amount of current itself is negligible, but the effects of having to build a system that generates it reliably on chip can potentially have dramatic effects on overall system power dissipation, PSRR etc. Here we have implicitly assumed that this is not going to occur, but no definitive proof can be claimed unless such system is developed and used to replace the ideal current source generating the $600\,nA$ bias current throughout our tests.

Closer examination of the waveform at steady state reveals the image shown in Figure 6.25. The computed average output voltage value at the output of the charge pump was equal to $2.312\,V$ whilst the ripples at the voltage reference and regulated output nodes had magnitudes of $876.5\,\mu V$ and $14.13\,mV$ respectively. Thus the power supply ripple limit has been exceeded. On the other hand, the regulated output voltage is sitting at around $1.55\,V$, which is within the acceptable range, but still significantly lower than the preferred value of $1.8\,V$.

The key performance metrics of the design can be summarised and shown in table 6.26.

Simulating with a photodiode macromodel as a power harvester element

Another useful experiment would be to replace the constant voltage VDD with a macromodel of an optical power harvester set at a certain constant current generation level. Then the ripples appearing on VDD and any variations in ring oscillator frequency could be studied. Similarly, a suitable photocurrent

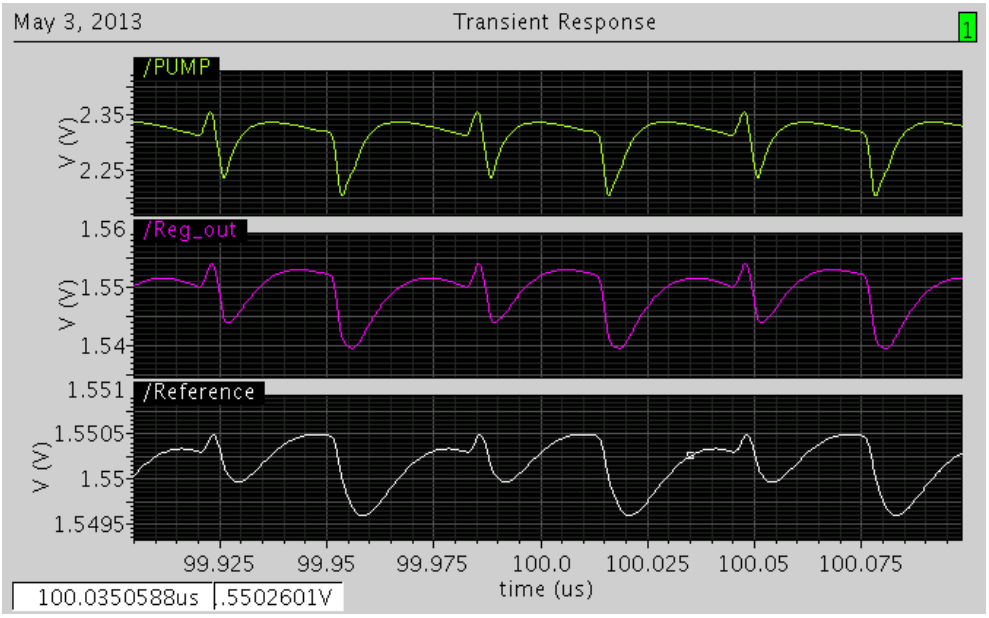

Figure 6.25: PMU steady-state behaviour showing voltage ripples. Voltage output traces at the output of the charge pump (green), the system output (regulated voltage node - pink trace) and the voltage reference output node (white).

Table 6.26: Key achieved specification for the PMU.

PMU PERFORMANCE SUMMARY			
Spec	**Performance**	**Units**	**Description**
12	≥ 12	μA	Current drive capability at load
1.5 - 1.9	1.55	V	Regulated output voltage
10	14.13	mV	Regulated voltage ripple

level can be found for which the PMU can operate properly and supply the necessary power to the load circuits. The value obtained would give some indications as to what light intensity is required in order to successfully operate the system.

The selected configuration for this test includes a theoretical solar harvester of the p-diffusion on N-well (pN) configuration measuring $1 \times 0.5\,mm$ and shaped as a rectangle[21]. As a result it features an area of $0.5\,mm^2$ and a perimeter of $3\,mm$. Using this element and a photocurrent source pushing $3\,mA$ we obtained a set of results. This is admittedly a very large amount of photocurrent for a diffusion-based photodiode but it was selected as a 'safe but realistic' choice i.e. potentially achievable under a powerful laser and fairly certain to succeed in operating the PMU and a $12\,\mu A$ load. The bias current sent to the Op-Amp is the same as in the standard ideal configuration. We shall call this new test configuration 'standard macromodel PMU configuration' or in short 'standard macro' configuration. If any parameter of this standard is then altered, it shall be explicitly mentioned.

Note: we have obtained photocurrent values of approx. $115\,\mu A$ in $300 \times 300\mu m$ devices on TED dies (TED-7-n3). These are n-diffusion on triple-well devices which are theoretically expected to have the most intensely doped regions (and therefore narrowest depletion regions) and are manufactured in a technological node with smaller feature sizes than the SVJ dies where the PMU resides. For these reasons we can expect these figures to be pessimistic estimates for photocurrents arising from a pN type junction on SVJ. The tests that gave rise to these figures were carried out with a powerful $70\,mW$ laser with $637\,nm$ central wavelength and unspecified wavelength spread (LP637-SF70 - ThorLabs). A very crude approximation therefore shows that if we adjust for device sizes, our theoretical $1 \times 0.5\,mm$ device may achieve at least 5.5 times the current output of the tested TED devices under the assumption that their average irradiance figures are the same. This already accounts for over $0.6\,mA$. Using even more powerful lasers and expanding the physical dimensions of the optical harvester may complement the difference. Real evidence though can only be found in the form of laboratory tests and real measurements. As such, the premises of this test remain rather speculative, yet the obtained results may prove educational.

Note: In the experiment above we need to consider of laser damage: Repeats ran with all our devices under the $70\,mW$ laser did not reveal any performance degradation over time, despite the long exposure times of the die to the radiation (over one hour to conclude an experimental 'round' covering all test devices on a single die - the die was exposed to continuous wave laser illumination throughout the 'round'). In reality, however, an experiment ran with the specific purpose of uncovering any laser damage should be ran to prove or disprove our implicit hypothesis that no significant laser damage was sustained by our dies.

A close-up of the system at steady state can be seen in Figure 6.26. The behaviour of the system from start-up to steady state can be seen in Figure 6.27. The results of the test with the photodiode macromodel are summarised in table 6.27.

The information gathered from this test leads to certain observations:

- The voltage supplied by the photoelectric element shows an approx. $10\,mV$ ripple but the clock signal seems to still feature a fairly balanced duty cycle. Thus, charge pump operation should not differ significantly between the 'realistic' set-up that includes the photoelectric element macromodel and the simplistic set-up that makes use of the ideal voltage source.

[21] The design repository can be found in appendix A for a library of possible designs.

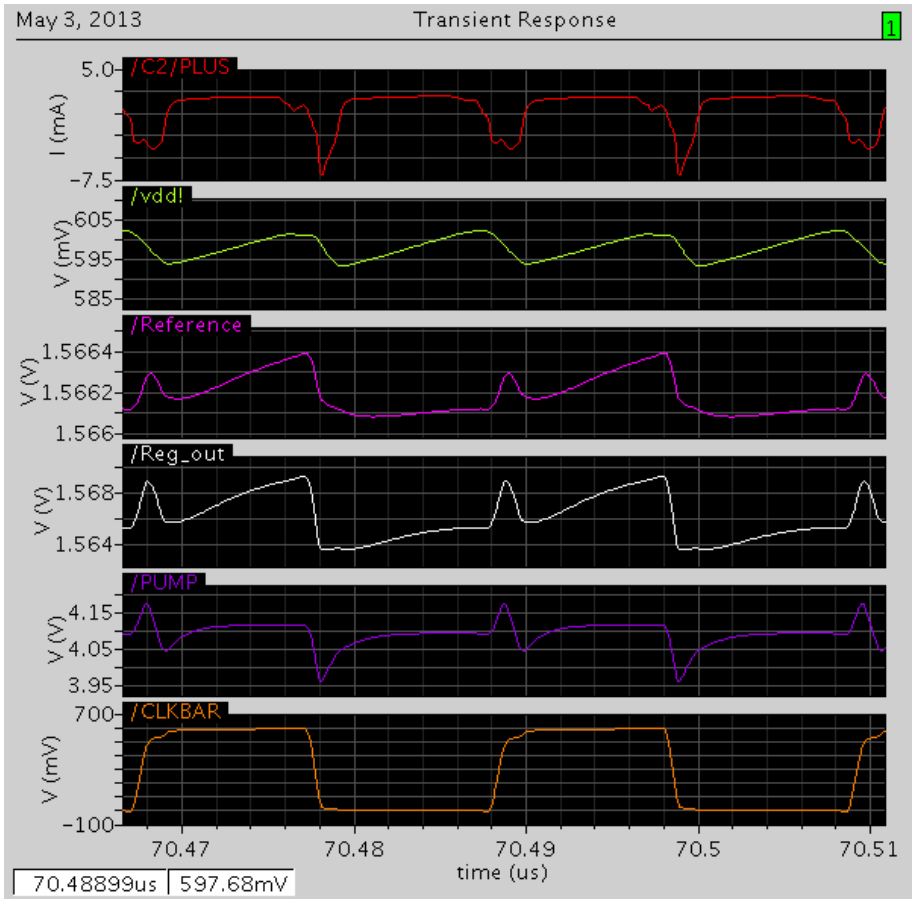

Figure 6.26: Sample transient trace of key signals at steady state for a full PMU test in the standard macro configuration. See text for test parameter set-up. Red trace: current drawn from the photoelectric element. Green trace: fluctuation of the voltage across the photoelectric element and therefore also of the VDD servicing the ring oscillator and clock generator modules. Pink trace: Voltage reference output. White trace: regulated output voltage. Purple trace: Charge pump output voltage. Orange trace: $\overline{CLK}$ signal serving the charge pump.

Table 6.27: PMU test result summary in the standard macro configuration. vdd! : voltage across photoelectric element. f_{clk}: clock frequency. V_{ref}: average voltage reference output voltage. V_{reg}: average regulator output voltage. V_{pump}: average voltage at the output node of the charge pump. $\overline{I_{dio}}$: average current dissipated by the diode component of the photoelectric element.

PMU TEST RESULT SUMMARY		
Parameter	Result	Units
vdd!	598.2	mV
f_{clk}	90.9	MHz
V_{ref}	1.566	V
V_{reg}	1.57	V
V_{pump}	4.09	V
$\overline{I_{dio}}$	657.6	μA

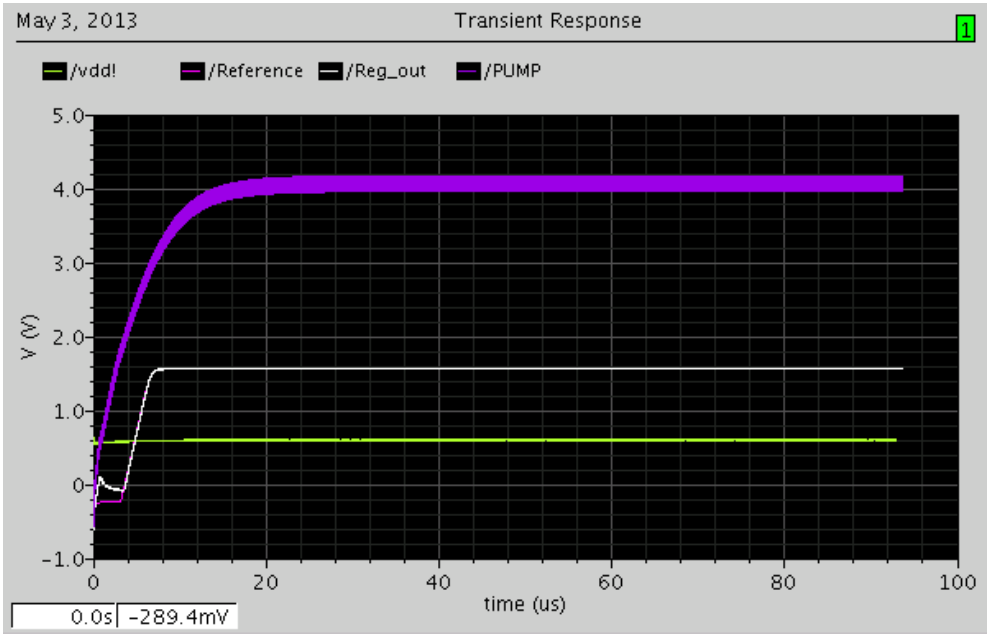

Figure 6.27: Overall transient behaviour of the PMU from start-up to steady state. Purple trace: charge pump output node voltage. White trace: regulated voltage output. Pink trace: voltage reference output (largely hidden behind white trace). Green trace: photovoltaic element output voltage.

- Overall ripple at the regulated output terminal seems reduced compared to the case with the ideal voltage source VDD (approx. $6\,mV$ vs. $14\,mV$). This may be because unlike in the case where our VDD is generated by an ideal voltage source, the photodiode macromodel simply isn't capable of providing 'spikes' of very high current at the exact moments when the clock generator is toggling states (see figure C.5). Instead, when current demand peaks, the macromodel VDD generator responds by experiencing a drop in its voltage, which in turn reduces the power requirements of the toggling clock generator. The overall effect is that the clock generator momentarily slows down during state transitions and therefore allows the charge pump to generate gentler ripples. Of course only a proper mathematical analysis can fully uncover the details, but this intuitive explanation is a reasonable starting point. Note: this reduced ripple now meets specs.

- The PMU seems to operate on a roughly $0.6\,V$ supply from the power scavenger and generate an approx $4.1\,V$ unregulated output voltage at the end of the charge pump cascade. This is clearly too far above specs and indicates that our system can operate even if the photocurrent level is substantially smaller than the $3\,mV$ imposed throughout this test session.

A few more test runs were also performed in the standard macro configuration, but with different photocurrent values. The basic metrics were collected in table 6.28.

We observe that even with as little as $1\,mA$ photocurrent the system can still provide the required regulated output voltage (approx. $1.5\,V$) for the $12\,\mu A$ constant current load. Moreover, as the voltage across the photoelectric element decreases due to reduced available photocurrent, the effective impedance relation between the diode component of the photoelectric element and the charge pump changes to the

Table 6.28: Important performance metrics for the standard macro test configuration, but with photocurrent treated as a parameter. I_{photo}: photocurrent. vdd!: voltage across photoelectric element. V_{pump}: average voltage at the output of the charge pump. V_{reg}: average regulated voltage level. $\overline{I_{dio}}$: average current dissipation by the diode element of the power harvester. Current loss: % of photocurrent wasted via the power scavenger diode element. All measurements are taken at steady state.

PMU VS PHOTOCURRENT					
I_{photo}	vdd!	V_{pump}	V_{reg}	$\overline{I_{dio}}$	Current loss
mA	V	V	V	μA	%
3	598.2	4.09	1.57	657.6	21.92
1.5	553.1	3.49	1.56	113.3	7.55
1	521.3	2.76	1.56	39.37	3.94

favour of the latter. The current wasted by crossing the forward biased diode becomes an increasingly small percentage of all available photocurrent.

The overall conclusion is that the PMU system shows itself as very likely capable of operating within specs in reality, provided that powerful light is shone on top of a sufficiently large photoelectric element. This, with powerful light sources that can be easily found on the market (e.g. the ThorLabs LP637-SF70 70 mW red laser used before) and photo-element size that is not unrealistically high for CMOS implementation.

Important note: Quite conspicuously, we have not calculated power efficiency values (power to load over power to PMU systems) for the PMU under various operating conditions. This is for a couple of reasons: a) Efficiency is highly dependent on operating conditions such as level of illumination (photocurrent magnitude) and load current. In fact, for each level of illumination that can sustain a load circuit (i.e. provides enough power to feed the PMU and leave some for the load circuit) there will be a specific maximum load current level that can be drawn sustainably over indefinite periods of system operation. Finding this value is not easy to calculate and too time-consuming to simulate. b) Given that under all circumstances mA worth of photocurrent lead to μA worth of load current (albeit at approx. 3× the voltage), we know that obtained efficiency rates would be extremely small and certainly well below what is achievable with slightly more complicated CMOS technologies. An example of this would be technologies that offer native devices with very low threshold voltages. Using such transistors could minimise threshold voltage-related losses in the charge pump (but at the cost of weaker OFF-resistances) and allow for the design of a faster (but more power-hungry) ring oscillator that could potentially force the charge pump to draw more charge from the photoelectric element at each step. Of course, the only way to find out the extent of such savings would be to test it on a simulated system, but we can expect such savings to exist.

Should the reader wish to calculate PMU efficiency figures they are given by the formula $\frac{I_{load} \cdot V_{reg}}{I_{photo} \cdot vdd!}$, or in words: load current times regulated voltage output over photocurrent times voltage across photoelectric cell. It is easy to realise that the numerator represents the power dissipation at the load whilst the denominator represents the power dissipation of the whole system. Most of the photocurrent will flow either down the forward-biased diode element inherent in the photoelectric cell or through digital circuitry keeping the charge pump running in proper fashion. Note: The component of the photocurrent that is earmarked for the load, however, will flow into a capacitor and raise the corresponding plate

potential from an ideal GND value to an ideal $vdd! - V_{th}$ value. Thus the charge going into the charge pump will lose, on average $\frac{vdd!+V_{th}}{2}$ instead of a full vdd!. This should be included in the calculations, but when system power dissipation is so strongly dominated by the PMU and the forward-biased diode of the power harvester the error is tolerable.

6.9 Conclusions

In this chapter we have set the basis for three very important functions that any integrated circuit that does not possess the luxury of bond-wires must perform in order to fulfill any useful function: a) Receive optical power and convert it to electrical power, b) convert the electrical power arriving from the photoelectric element into a useful power supply; stable and providing enough headroom for circuitry to operate and c) receive optical signals and convert them to electrical signals.

Objectives a) and c) are both achievable by using the exact same structures: pn-junctions that act as power harvester elements. Indeed the similarities between photoelectric elements acting as power scavengers and signal capture devices are far more than their differences. They will operate at similar open-circuit voltages (important for triggering digital circuits) and be subject to the same layout constraints but they will differ in scale. Power harvester elements will cover much larger areas than 'optical ports'. Thus, by studying how layout, junction type choice and technology choice affect power harvester performance we have solved two problems at once: understanding how the engineer's design choices can deliver better elements that can both power the integrated circuit and feed it with input signals in a completely contact-less way.

Objective b), on the other hand, involved circuit design where we have shown that a 'leisurely' designed circuit (i.e. not aggressively optimised for power conversion efficiency) can use the apparently meager power budget that our power scavenger elements provide it in order to operate a stable, $1.5\,V$ power supply and be capable of maintaining a load that consumes $12\,\mu A$ constant current, all without straining the power harvesting elements. It is expected that using slightly more elaborate CMOS technologies that allow, for example, the use of native devices with very low threshold voltages, may allow engineers to design PMU systems that can operate with far lower power dissipation than our system currently does.

Whilst tackling photoelectric elements we have studied the effects of inter-die variation on their performance and then proceeded to extract areal and side-wall power coefficients that describe the contribution to power output of each square micron of areal junction component and each micron of side-wall junction component. Throughout the procedure we discovered that though the idea of splitting junctions into areal and side-wall components according to their nominal layout areas and perimeters does seem to provide reasonable results for larger junctions and well-based junctions, the model shows signs of breakdown when considering smaller and diffusion-based junctions. Thus, the numbers obtained for power coefficients, useful indicators of performance though they may be, should be taken with precautions and a new attempt should be made to understand exactly how areal and side-wall junction regions interact. This would imply finding a new model, perhaps by redefining the concepts of areal and side-wall junction components, and lies in the domain of future work.

Power coefficients were obtained for most junction types available in the technologies under study

and, flawed though they may be, have provided some insight into how different components of different junctions in different technologies operate. NS type junctions were found to be the optimum devices for designing photoelectric elements in all technologies whilst other junction types compete for second place and below in a manner that depends on the specific manufacturing technology.

Finally, we considered how the key learnings obtained fit into a design environment with realistic constraints, particularly with regards to layout area and available junction types. *Junction type considerations:* Sadly, the stellar performers -N-well on substrate (NS) junctions- are not available for use as power harvesters because their p-side terminal is by definition shorted to substrate/GND. Thus, this leaves triple well on N-well (3N), p-diffusion on N-well (pN) and n-diffusion on triple well (n3) devices still competing for the best practical choices for the role of power providers. We found out that the best out of these options is likely to strongly depend on the specific characteristics of each technology. *Layout area considerations:* By analysing our designs in terms of areal and side-wall junction densities (extent of each component per unit area of device footprint) we attempted to extract some further learnings. The matter at hand requires much further study and very careful consideration of layout geometries but the figures obtained so far tend to show that a balanced mix of areal and side-wall components performs rather well. These results are not absolutely conclusive, but nevertheless serve as useful hints for good design.

Bibliography

[1] S. M. Sze, *Physics of Semiconductor Devices*. Wiley, 2nd ed., 1981.

[2] R. Wood, "Laser induced damage thresholds and laser safety levels. do the units of measurement matter?," *Optics & Laser Technology*, vol. 29, no. 8, pp. 517 – 522, 1998.

[3] F. Horiguchi, "Integration of series-connected on-chip solar battery in a triple-well CMOS LSI," *Electron Devices, IEEE Transactions on*, vol. 59, pp. 1580 –1584, june 2012.

[4] C.-M. Chang, C.-L. Chang, and C.-C. Chang, "Synthesis and optical properties of soluble polyimide/titania hybrid thin films," *Macromolecular Materials and Engineering*, vol. 291, no. 12, pp. 1521–1528, 2006.

[5] C. H. Andreea Irina Barzic, Iuliana Stoica, *High Performance Polymers - Polyimides Based - From Chemistry to Applications*. Dec. 2012-12-19.

[6] R. Pelliconi, D. Iezzi, A. Baroni, M. Pasotti, and P. Rolandi, "Power efficient charge pump in deep submicron standard CMOS technology," *Solid-State Circuits, IEEE Journal of*, vol. 38, pp. 1068 – 1071, june 2003.

[7] R. Pelliconi, "Us patent 6,819,162." patent, September 2004.

[8] R. Baker, CMOS *circuit design, layout, and simulation*. Hoboken, John Wiley and Sons, Inc., 3 ed., 2011.

Chapter 7

Conclusions and future work

The final chapter of this thesis is dedicated to the following three objectives: a) Summarising the project from concept to final results and interpretation, b) illustrating the original contributions in this work and c) mentioning the potential for applications and the vision behind this project and offering recommendations about how the research shown here can be progressed to the next stage. Finally, an extra section detailing the opinion of the author about the entire project in a more informal way forms the end to the thesis.

7.1 Project summary

This project was started with the aim of developing the necessary electro-optical structures and basic associated circuitry that would allow an integrated circuit die to operate and communicate with its environment without the use of bond wires. On top of this objective we imposed the constraint that all resulting systems would have to be manufacturable in standard, commercially available CMOS technologies.

Over the course of the project we have determined that in order to achieve our aims under the given constraints we would need to develop structures that carry out the following tasks: a) Data read-in. b) Data read-out. c) Power recovery. d) Power management. Tasks a-c were achieved by using electro-optical devices whilst task d was handled by dedicated circuitry. This created a natural division of tasks between the 'devices-oriented' group and the 'circuitry-oriented' group.

The 'devices-oriented' group of tasks started from a physics basis. We observed that all three tasks in the category (data in, data out[1], power in) can be carried out by simple, pn-junctions. Data read-in and power recovery thus relied on the exploitation of the inter-band absorption phenomenon whilst data read-out relied on the exploitation of the free carrier absorption phenomenon; a phenomenon that allows a standard pn-junction to act as an electro-optical modulator.

Given this situation we decided that we would create a single set of pn-junction-based devices and test all of them for both electro-optical modulation and for optical power recovery. The set of test devices was conceived and designed so that it would allow us to assess performance in both modulator and

[1] By extension and extrapolation from power in - if one can be done, so can the other via a variety of well-known methods such as power supply, non-return to zero duty cycle ratio modulation etc.

power harvester regimes. Resulting data would then be processed in order to allow us to understand how junction geometry and type, as well as the selection of CMOS technological node affect affect respective performances.

The 'circuitry-oriented' task involved using circuit design in order to create a power management unit that can convert the typically low output voltage provided by optical power scavengers to a stable DC power supply with enough headroom in order to run large varieties of load circuits (in our case $1.5\,V$). The power management unit was fully simulated, laid out and fabricated.

The measured results emanating from our electro-optical devices showed us that further study is required in order to understand how junction geometry and type, and manufacturing technology affect electro-optical modulation. In terms of power recovery the results showed that well-based junctions tend to out-perform diffusion-based junctions in all technologies and that technologies with finer feature sizes tend to under-perform their coarser counterparts. However, the specifics of each manufacturing process also affect performance very significantly which means that the best way to be certain of power recovery performance in a given technology is to run tests on dies of that specific technology. Finally, simulated results from our power management unit in combination with power recovery results showed us that if a reasonably powerful light source is available, then powering load circuits consuming up to $12\,\mu A$ should not pose significant problems or require excessive amounts of chip area dedicated to power harvesting.

7.2 Original contributions

The contributions of this thesis to the field of electro-optical communications can be summarised thus:

- Electro-optical theory: One contribution in this thesis arose from studying doping profiles arising from idealised versions of ion injection processes typically used in CMOS fabrication (planar diffusion and ideal ion implantation) and considering their predicted effects on modulator and power harvester performance. We then examined the doping concentration vs. location maps they create under different assumed manufacturing conditions. Finally, we commented on the gradients of the aforementioned maps and explained how the magnitudes of the gradients can influence both modulator and power harvesting performance.

  Furthermore we have developed a new, so-called 'beam fascicle analysis' version of finite element analysis in order to provide a useful means of studying electro-optical modulation in CMOS-fabricated pn-junctions. The principle was applied to three specific, 'canonical' cases, distinguished by the geometrical configuration of the emitter-modulator-receiver assembly. We call the cases 'canonical' because they make use of extensive symmetries and ideal assumptions in order to illustrate the principle with an eye towards following up with numerical methods in order to tackle more complex cases.

- Electro-optical modulator implementation: Throughout this project a number of practical modulator designs in different technologies have been developed and presented. These designs, along with our recommended changes can be implemented as part of future projects. Furthermore, they were tested and assessed and despite the inconclusive nature of our results the entire procedure forms a significant stepping stone towards achieving the goals of the project.

Through the long development of novel test benches capable of measuring electro-optical modulation we have managed to identify key issues, notably sources of uncertainty. When tackling a phenomenon as weak and delicate as free-carrier absorption, the design of the test-bench is absolutely crucial. We have thus summarised all our observations towards the creation of a suitable test-bench, including requirements placed upon the test integrated circuit itself.

- Optical power recovery: We have carried out an extensive study of how technology and geometry affect the performance of both electro-optical modulators and power harvesters whilst also considering variations between dies. Results obtained from our pn-junction structures on the topic of power recovery showed clear differences between device types and technologies. Results have been interpreted on the basis of a model whereby areal and side-wall junction components are considered separately and have also been discussed (e.g. in 7.1). Our basic model can be expanded as part of future work, but already has provided us with data on whether to stress area or side-wall prevalence during optical power recovery device layout. In fact, the tables we compiled, showing normalised power output per unit layout area and unit layout perimeter at the maximum power delivery point as a function of technology and junction type, are a very useful resource for anyone trying to design efficient and compact power recovery structures and thus they form a key contribution of this project.

- Other: We have developed a novel packaging technique that allows CMOS dies to be incorporated into a standard, plastic package casing while at the same time allowing beams of at least visible and infra-red light to cross through the entire 'chip & package' assembly. This was achieved by drilling a hole in the package, mounting a thin glass lid on top of the hole, mounting the die on top of the lid, bonding the die and finally sealing it with yet another glass lid covering the chip, the bond wires and the package pads. The assembly provides an elegant way of accessing the chip optically while at the same time protecting everything inside the 'sealed chamber' from the ingress of dust or other contaminants. The technique is also suitable for cases where the die to be mounted and bonded is smaller in dimensions that the hole in the package (i.e. the die could slip out of the package through the hole). We have no reason to believe that this packaging technique cannot be automated and improved to industrial standards of quality and through-put; an advance that can potentially render the testing of special opto-electronic devices significantly easier.

7.3 Applications and future work

Our research was developed with the aim of eventually applying this technology to lab-on-chip (LOC) systems that require IC dies to come into close contact with fluids. On a system-level scale, this technology would revolve around special, planar 'receptor boards' that can: a) host LOC ICs in specifically designed notches, fixed in place by suction and b) feature in-built IR emitters and receivers for enabling data read-out from the IC. This would be a very easy to use and convenient way of packaging lab-on-chip dies, offering the ability to quickly change dies without having to destroy large parts of the set-up in the process. Moreover the chip-board assemblies would be inherently planar. The IR emitter/receiver pairs could be of the order of hundreds of microns and thus much larger than the feature sizes used

in semiconductor fabrication while at the same time reasonably easy to manufacture and affix to the 'notch' designed to host the IC. This configuration may not allow for high chip-to-board interconnect density, but for generally low-bandwidth applications may prove to be more than sufficient. Developing the board-side of the system could be an interesting topic for future research.

On the chip-side of the system, future work can concentrate on building up from where this work has reached. We note a number of improvements and advancements to our test set-ups that can be implemented in future projects:

- Reducing pick-up: Drive the signal from an on-chip driver, optimised for low power consumption.
- Improving intrinsic modulation efficiency: Use longer wavelengths by employing mid-infrared light sources and photodetectors.
- Reducing noise: Pick inherently low noise and if possible also thermally stabilised light source assemblies. Keep away from sources of electrical noise. Use low noise pre-amplifiers before feeding the received signal to any other instrumentation. Carefully craft PCBs and attempt to hide signal-carrying wires within e.g. ground planes.
- Improving productivity: Develop and use industrial-grade, highly automated test platforms in order to carry out otherwise menial tasks such as connecting DUTs to test ports and aligning for maximum photocurrent.

The benefits of achieving a working, contactless power/data transfer system would be substantial with regards to solving packaging issues often encountered by researchers who wish to mix the inherently hostile worlds of electronics and fluidics together. Therefore, we conclude this thesis by expressing our belief that this technology is (either unmodified or with the EM-based output modification) potentially a practical solution to the issue of packaging dies intended for close contact with conductive liquids; a solution that merits further study.

7.4 Concluding remarks

This section is dedicated to a more informal collection of the author's remarks on the subject of the overall project. Specifically, a point on the practical implementability of a contact-less chip-to-board communications system has to be made.

In the opinion of the author, the idea of a contact-less chip-to-board communications system is useful and sound, and the achievement of power harvesting and signal read-in via optical coupling and the use of photo-electric elements is easily implementable and practical. On the other hand, the implementation of electro-optical modulation for the purposes of data read-out is ridden with problems.

To begin with, we consider the issue of power harvesting. The literature review of this thesis does include numerous examples of inductively coupled systems that are geared towards transmitting power to the system (e.g. [1, 2] etc.), but they tend to rely on post-processed coils (although not exclusively - see [3]), which renders large scale manufacturing much more complicated than its optical power harvester

counterpart. In terms of performance it is true that inductive systems can offer mW worth of power, but if the system to be powered is an energetically frugal signal processor (as opposed to e.g. a power-hungry stimulator) the μW level powers that optical systems can provide with great ease should be more than sufficient. Note, however, that the hidden cost of optical power harvesting is in front-end real estate (as opposed to inductors, which are always either in the back-end or post-processed). In conclusion: no clear winner.

In terms of data read-in both inductive and optical systems share the capability of shifting data into the system by modulating the power supply in a classical non-return to zero duty-cycle rationing way[2]. Furthermore, in both cases an auxiliary structure can be built that receives the data independently of the power harvesting structures. In the optical case, a small (let's say 10×10 micron) photodiode operating completely independently of the large power harvester can act as a data receiver and 'talk' to dedicated circuitry. For an example where a specialised data receiver coil is used see [4]. Once again: no clear winner.

Finally, the really thorny issue of data read-out is far more difficult to tackle with electro-optical means. The electro-optical solution is based on a phenomenon that we have shown to be extremely weak; weak to the point where very specialised circuits will be required to even attempt to extract data from it. We do believe that with very careful engineering a system that exploits the phenomenon of free carrier absorption may eventually become practical, but the need for such specialised circuitry means that the optical method can no longer be shown to clearly 'win' over the inductive methods. Therefore, for more commercial applications it would seem that the old and proven inductive methods might work better. As a very simplistic example, a receiver coil could be connected to a rectifier that is switched on or off by an on-chip circuit. The rectifier then feeds an emitter coil. When the rectifier is on, a lot of the power entering through the receiver coil will be shuttled on to the emitter coil at double the main frequency and that should, at least in theory, be easily detectable off-chip. When the rectifier is off, the emitter coil remains silent and the second harmonic of the input frequency detected off-chip should dramatically diminish. This system seems straightforward in theory, but the author has not attempted to implement an inductive-based system, so we may only speculate on the practical difficulties associated with its construction and operation, but given the difficulties encountered whilst building the optical system, we believe we can declare the inductive methods as the clear winner of this 'round'.

Thus, the conclusion of the above remarks from the point of view of the author is that the tremendous difficulties associated with the sheer weakness of electro-optical modulation in CMOS makes the implementation of an inductive-based data read-out system imperative. However, the data read-in and power harvesting parts of the system should not necessarily follow suit; they can be easily implemented optically. Ultimately, the nature of the application shall determine what the best solution is in each case: how precious is the front-end real estate? how damaging is crosstalk for our system? is the option of using post-processing for building coils available? All these and many more are questions that will determine what parts of the system operate under what regime, but one thing is almost certain beyond a shadow of doubt: the data read-out will have to be done inductively, at least until the optical method can catch up and if indeed it ever does catch up in terms of performance and simplicity.

Thus we can end by envisaging a whole array of possible 'hybrid' systems where power recovery and

[2]Power supply spends a certain amount of time x in the 'high power transmission' state and a certain amount y in the 'low power transmission' state and the transmission of a '0' or a '1' is determined by the ratio x/y.

data read-in is handled either by optical or inductive coupling means on the basis of what works best for each application whilst data read-out is relegated to an inductive structure. Yet enthusiasm should not ignore the fact that this class of solutions has a host of its own problems: the use of coils cannot be avoided and the system will be requiring a lot of power for data read-out, although in theory the off-chip receiver should be able to recover part of it. Ultimately, further study and development is needed in order to solve the conundrum of creating simple, reliable and power-frugal contact-less CMOS integrated circuits.

Bibliography

[1] D. Dudenbostel, K.-L. Krieger, C. Candler, and R. Laur, "A new passive CMOS telemetry chip to receive power and transmit data for a wide range of sensor applications," in *Solid State Sensors and Actuators, 1997. TRANSDUCERS '97 Chicago., 1997 International Conference on*, vol. 2, pp. 995 –998 vol.2, jun 1997.

[2] J. Von Arx and K. Najafi, "A wireless single-chip telemetry-powered neural stimulation system," in *Solid-State Circuits Conference, 1999. Digest of Technical Papers. ISSCC. 1999 IEEE International*, pp. 214 –215, 1999.

[3] A. Shameli, A. Safarian, A. Rofougaran, M. Rofougaran, and F. De Flaviis, "An rfid system with fully integrated transponder," in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2007 IEEE*, pp. 285 –288, june 2007.

[4] Y. Y. et al., "Simultaneous 6-gb/s data ad 10-mw power transmission using nested clover coils for non-contact memory card," in *IEEE journal of solid-state circuits*, vol. 47, October 2012.

Appendices

Appendix A

CMOS electro-optical modulator/photo-transducer design repository

We shall use this section of the appendix to describe the entire design procedure of all modulator devices that have generated results throughout this research. The objectives that each device had to fulfill will be laid out along with the constraints imposed by the specifications of this research before an analysis of the design considerations that stem from the said objectives and constraints is carried out.

We found that under the objective of good NIR modulation and the restriction of using nothing but commercially available CMOS technologies the available levers that an engineer has at their disposal consist of the technology selection, the pn-junction type selection and the device geometry. Based on this observation we have decided to design 29 test devices representing three different CMOS technologies, featuring unique, individual geometries and each consisting of one or more pn-junctions from a variety of pn-junction 'types'.

Throughout this section an overview of each die will be given (technology properties, overall floor-plan, diagrams of their structure and microphotographs from the resulting, fabricated IC dies.) and subsequently each device will be described and illustrated both in terms of top-view and as a cross-section. Important note: Tables with key device information are provided for each device. This includes calculated measures of the illumination-exposed areas and perimeters of each device. These area and perimeter values have all been computed with those areas and perimeter segments that are shaded by any metallisation (including the metal tracks that are used to connect the various junctions to their respective bond pads) taken into account.

A.1 The 'Ninja' design family: modulators designed in 0.35 micron technology

The $2.5 \times 2.5 \times 0.53\,mm$ 'Ninja' was the first device to be fabricated in order to test optical modulator structures and provide the proof of concept. It was implemented in the AMS 0.35 micron AMS35 technology. The technology offered the conventional NW/sub, p+/NW and n+/sub junctions with the

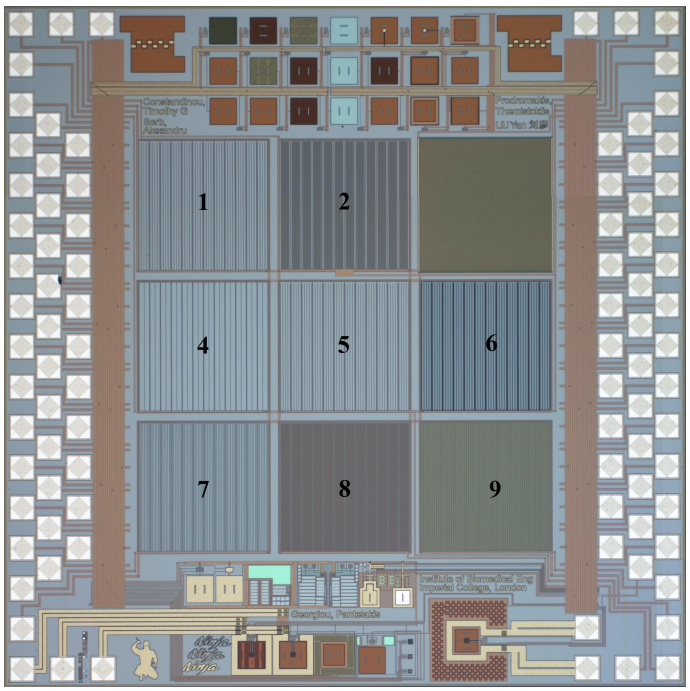

Figure A.1: Ninja device microphotograph with all optical modulator devices labeled by their identification numbers. The numbers correspond to their unique identifier numbers. The device that would normally feature the identifier '3' was a non-optical structure used for other purposes.

addition of the so-called 'butting diffusion' junction where a p+ diffusion region abuts an n+ region. All these junction types were represented in the eight optical devices residing on the Ninja. Each of these devices had a $479 \times 479\,\mu m$ footprint. A labeled microphotograph of a complete Ninja die can be seen in Figure A.1 whilst the basic properties of the Ninja design family can be seen in table A.1.

Table A.1: Summary of the basic features of the Ninja die family.

Ninja summary	
Technology	AMS35
Feature size	$350\,nm$
Size	$2.5 \times 2.5\,mm$
Die thickness	$530\,\mu m$
Modulator device number	8
Devices size	$479 \times 479\,\mu m$

The devices residing upon Ninja are the following (sorted by identifier number):

- NIN-1: NW/sub-type device featuring strips of N-well on substrate. Designed in order to provide for long side-walls.
- NIN-2: Nested junction device featuring p+/NW and NW/sub junctions.
- NIN-4: n+/sub device. Split into large areas of n-type diffusion over substrate it is dominated by the areal n+/sub junction.
- NIN-5: Large NW/sub squares where the areal NW/sub dominates.

- NIN-6: NW/sub device geometrically same as NIN-5 but with no passivation layer over its active area. Designed to try and discern the effects of the passivation layer.
- NIN-7: Butting diffusion device (p+ against n+ -p+/n+-).
- NIN-8: p+/n+ device similar to to NIN-7 but with a different geometrical configuration.
- NIN-9: Special n+/sub device where parts of the junction area have been covered by poly-silicon. Designed to study the effects of the polysilicon layer.

Each device will be presented with a layout diagram. The nomenclature for all devices within a given die is the same and for the NIN design family it can be seen in table A.2. Please refer to this table as a legend for all layout and cross-section view images of Ninja devices.

Table A.2: Legend to all layout and cross-section view images of devices residing upon the Ninja die. Out-diffusion regions mark the practical extent (as opposed to the design extent) of diffusion regions.

NINJA IMAGE LEGEND		
	Item	**Explanation**
Layout view	Stripy blue	Metal layer 1 (M1)
	Stripy white	Metal layer 2 (M2)
	Turqoise square	M1 to front end contact
	Pink square	M1 to M2 contact
	White/yellow boxes with dots inside	n+/p+ out-diffusion regions
	White outline box	N-well
	Green (webbed)	Diffusion regions (can be p+ or n+)
	Red (webbed)	Poly-silicon
Cross-section view	Mx	X-th metallisation layer
	Via	Connection between metallisation layers or metallisation and substrate
	n+/p+	n+/p+ diffusion regions
	NW	N-well
	3W	Triple well (p-type)
	Poly	Poly-silicon

NIN-1

NIN-1 (Ninja design family, device identifier no.1) is a NW/sub type device that consists of 1800 small 15×3 micron rectangles of well regions over substrate. The small size of each N-well element gives side-wall regions a prominent position, which allows the examination of the contributions to modulation of the side-wall component when overall device performance is compared to that of NIN-5 for example. The prominence of side-wall regions becomes evident given that $62280\,\mu m$ of perimeter enclose an areal N-well junction of $75204\,\mu m^2$. By comparison if the entire area covered by N-well areal junctions formed a single square the perimeter of the said square would be slightly over 554 microns. A layout view of a small section of the array of basic elements is shown in figure A.2 whilst the vital design information relating to NIN-1 can be seen in table A.3.

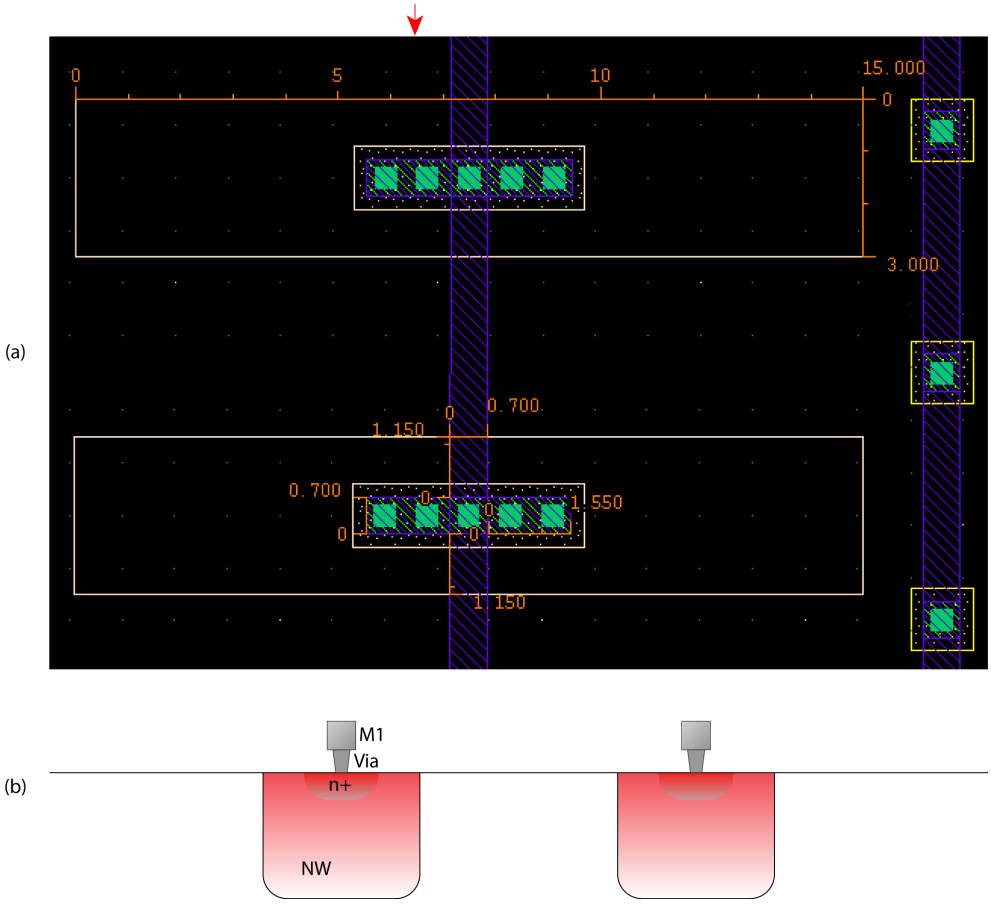

Figure A.2: Basic cell of device NIN-1 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) Measurements in microns are included (orange bars and numbers).

Table A.3: Basic features of device NIN-1

NIN-1 properties table	
Footprint	$479 \times 479\,\mu m$
Basic cell area	$41.78\,\mu m^2$
Basic cell perimeter	$34.6\,\mu m$
No. of basic cells	1800
Total design area	$75204\,\mu m^2$
Total design perimeter	$62280\,\mu m$

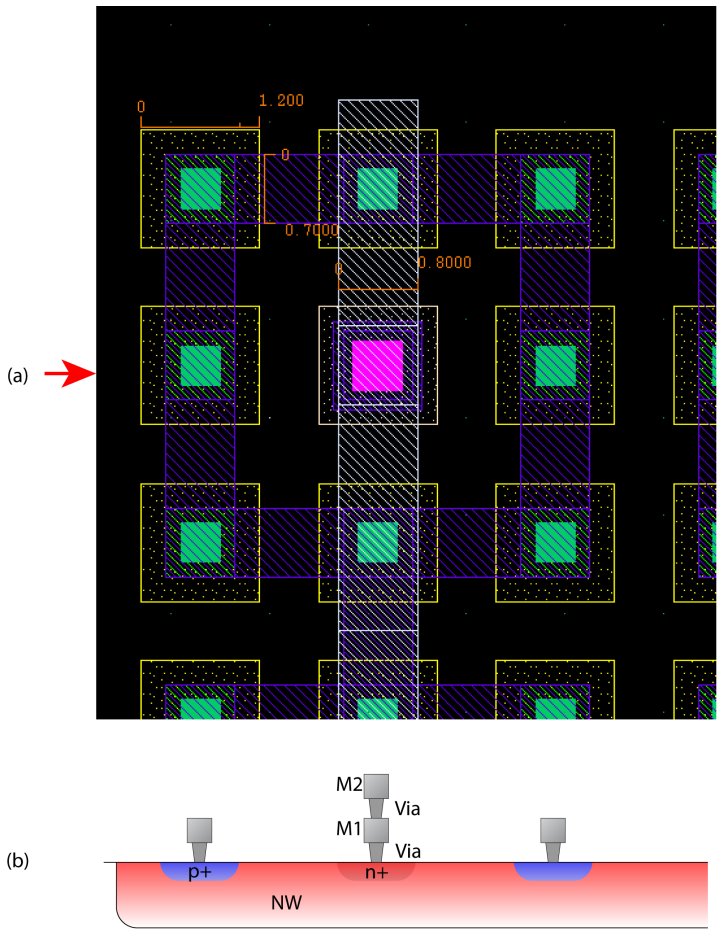

Figure A.3: Basic cells of device NIN-2 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) A metal to front end contact is hidden beneath the via as is betrayed by the out-diffusion region surrounding it. Important geometrical measurements are displayed in orange.

NIN-2

NIN-2 (Ninja design family, device identifier no.2) is a p+/NW/sub type device that consists of 12 large N-well regions that host a combined 6336 p+ areas. N-well regions measure $37 \times 479\,\mu m$ and were designed with such large dimensions primarily in order to act as a substrate for the p+ regions. The N-well regions feature a total area of approximately $84142\,\mu m^2$, fenced by $11500\,\mu m$ of perimeter. The p+ diffusion regions, on the other hand are arranged in small $1.5 \times 1.5\,\mu m$ islets with the intention of creating very large perimeters. Overall, p+ regions span approximately $27878\,\mu m^2$, fringed by a total perimeter of $152064\,\mu m$.

This comparably large perimeter area allows use of the device in order to analyse the behaviour of side-wall p+/NW junctions with little interference from areal components. Importantly, because of the small size of each individual p+ diffusion islet the areal component of each such islet can potentially act in a very different fashion to areal components of greater extent because of the proximity of each point within the islet to a junction edge. A layout view of a small section of the array of basic elements in shown in figure A.3 whilst the vital design information relating to NIN-2 can be seen in table A.4.

Table A.4: Basic features of device NIN-2

NIN-2 properties table	
Footprint	$479 \times 479\,\mu m$
Basic cell area (NW)	$7011.84\,\mu m^2$
Basic cell perimeter (NW)	$958.40\,\mu m$
No. of basic cells (NW)	12
Total design area (NW)	$84142.08\,\mu m^2$
Total design perimeter (NW)	$11500.00\,\mu m$
Basic cell area (p+)	$4.40\,\mu m^2$
Basic cell perimeter (p+)	$24.00\,\mu m$
No. of basic cells (p+)	6336
Total design area (p+)	$27878.4\,\mu m^2$
Total design perimeter (p+)	$152064.00\,\mu m$

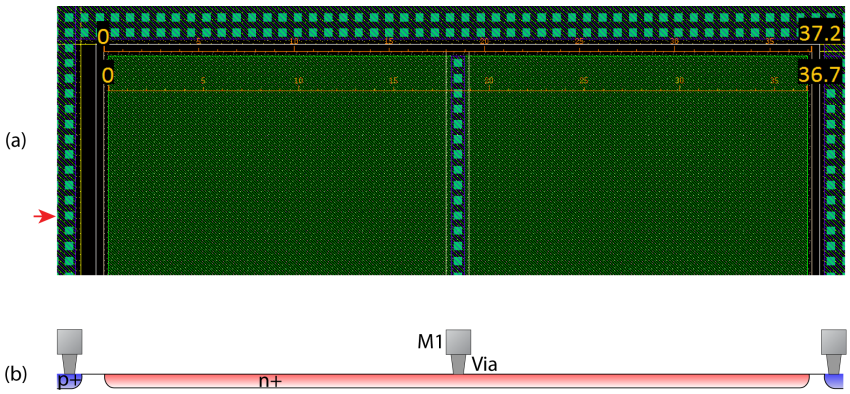

Figure A.4: Basic cell of device NIN-4 in (a) top-view and (b) cross-section. (a) The widths of the nominal diffusion and the out-diffusion regions of the basic cell are shown in orange. The contact 'frame' encircling the diffusion region from three sides is a guard ring. (b) Cross-section along the line indicated by the red arrow in (a). Part of the guard ring structures is visible.

NIN-4

NIN-4 (Ninja design family, device identifier no.4) is an n+/sub type device that consists of 12 large diffusion regions. Each region measures $478.3 \times 36.7\,\mu m$. The total area covered by these 12 regions measures $206625.6\,\mu m^2$ and is fenced by a $12342\,\mu m$ perimeter. The intended use of this design is to create an n+/sub type device exhibiting extended areal junctions with little interference from fringe regions. When results from NIN-4 are combined with results from devices with much more prominent side-wall components (e.g. NIN-2) the effect of the areal component of NIN-4 can be theoretically at least partially determined. A layout of a section of a basic cell can be seen in figure A.4 whilst important design information is summarised in table A.5.

NIN-5 and NIN-6

NIN-5 and NIN-6 (Ninja design family, device identifiers no.5 and no.6) are devices that share geometry entirely but differ in that NIN-5 is covered by passivation as nearly every other device is, but NIN-6 is not. The motivation behind this design decision was to determine to what degree the passivation layer

Table A.5: Basic features of device NIN-4

NIN-3 properties table	
Footprint	$479 \times 479\,\mu m$
Basic cell area	$17218.8\,\mu m^2$
Basic cell perimeter	$1028.5\,\mu m$
No. of basic cells	12
Total design area	$206625.6\,\mu m^2$
Total design perimeter	$12342\,\mu m$

affects performance. Geometrically these devices consist of 624 $15 \times 15\,\mu m$ N-well squares covering a total area of $133848\,\mu m^2$ encircled by $36566.4\,\mu m$ of perimeter. This configuration was chosen because it yields significantly more area for considerably less perimeter than for the NW/sub device NIN-1 and thus allows the study of larger areal NW/sub junctions. A layout view of a small array of basic cells is illustrated in Figure A.5 whilst key design information is displayed in table A.6.

Table A.6: Basic features of devices NIN-5 and NIN-6

NIN-5 and NIN-6 properties table	
Footprint	$479 \times 479\,\mu m$
Basic cell area	$214.5\,\mu m^2$
Basic cell perimeter	$58.6\,\mu m$
No. of basic cells	624
Total design area	$133848\,\mu m^2$
Total design perimeter	$36566.4\,\mu m$

NIN-7

NIN-7 (Ninja design family, device identifier no.7) is a device featuring the so-called 'butting diffusion' junctions where p+ and n+ diffusion regions are created in very close proximity and should theoretically give rise to pn-junctions with very high 'steady state' or 'bulk'[1] doping concentrations either side of the metallurgical surface. The specifics will depend on the exact manufacturing process, but the fact that the layout design rules state that butting diffusion junctions are created by designing diffusion region rectangles adjoining each other, would indicate that the two diffusion regions are likely to intermingle strongly. In practice this means that their individual doping concentration vs. location along the surface of the die functions will overlap strongly and create metallurgical surfaces at generally high doping concentrations compared to other junction types. Under the strict condition that the net doing concentration gradient is similarly very high (manufacturing process-dependent), a much better approximation to an abrupt junction can be created using butting p+/n+ techniques than other, more 'traditional' CMOS junctions. The risk, however, is that the high doping concentrations involved would force the depletion region across the pn-junction to become so narrow that tunneling effects would dominate and the rectifying nature of the junction would be largely lost. In that case the depletion region will also become 'contaminated' by free carriers therefore no longer operating as an ideal depletion region, which for our purposes means 'devoid of free carriers'.

[1] By which term one is to understand the doping concentration 'far enough' from the junction, in a region where the gradient of doping concentration vs. location is relatively close to zero.

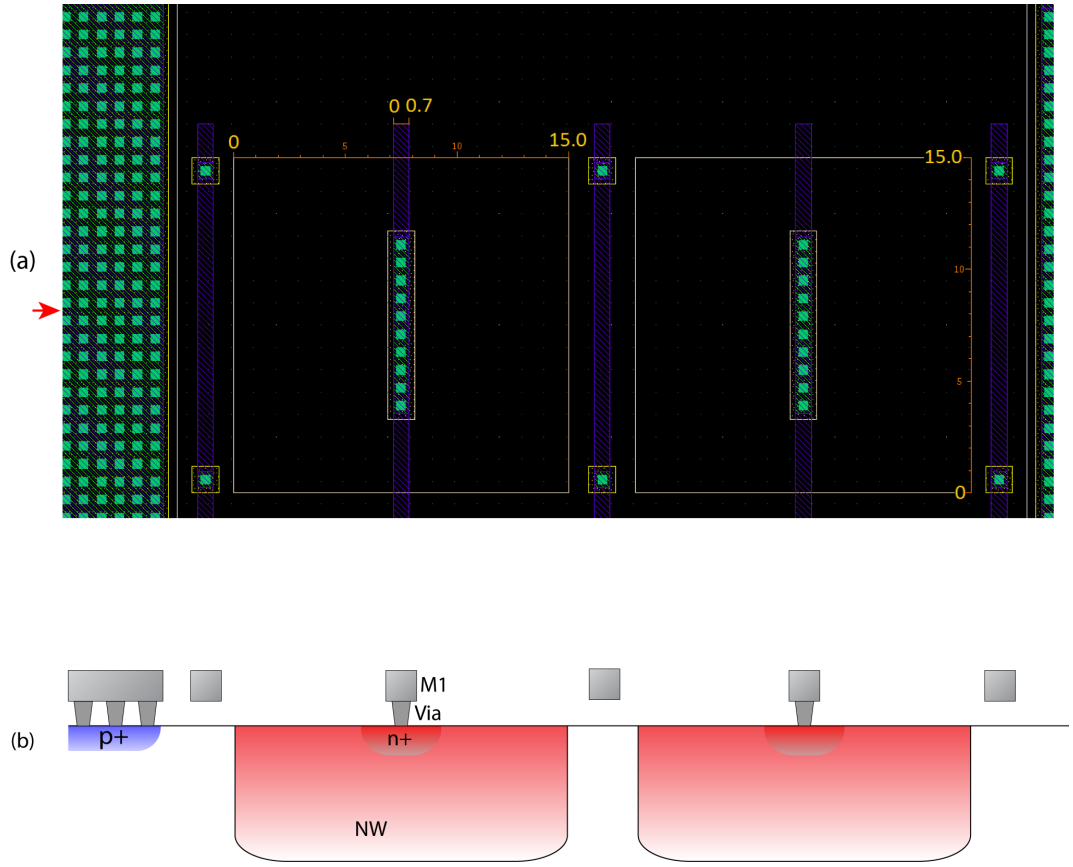

Figure A.5: Basic cell of device NIN-5 in (a) top-view and (b) cross-section. (a) Some measurements are also shown in orange (units of microns). (b) Cross-section of the devices along the line indicated by the red arrow in (a). Some of the guard ring structures are visible. The apparently floating chunks of M1 are the tracks that give off substrate anchor contacts above and below the cross-section line. Key metrics are shown in orange (units of microns).

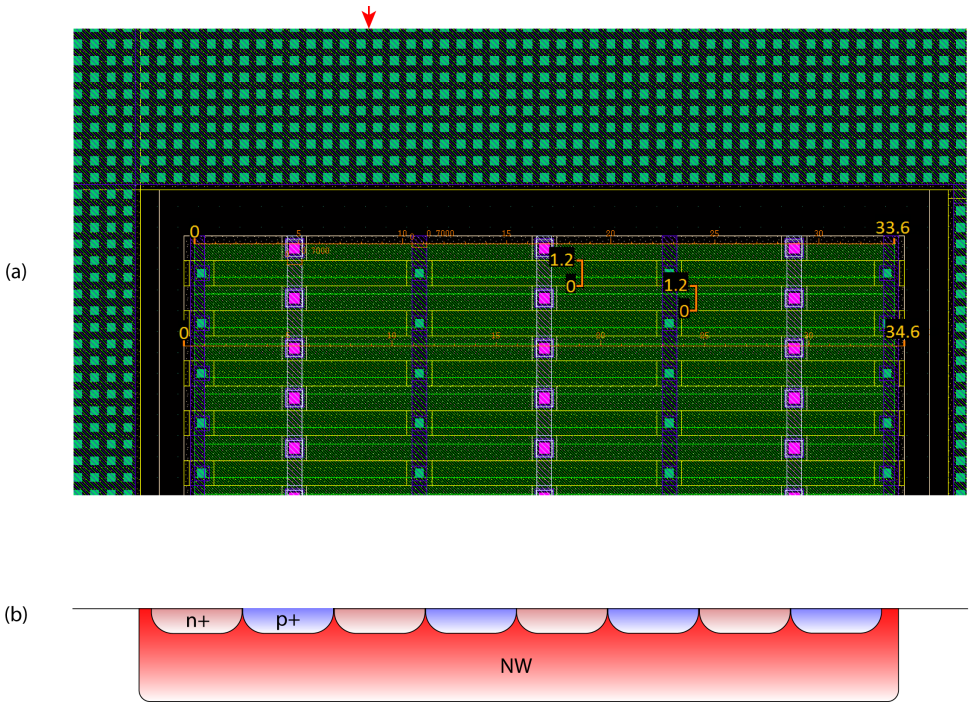

Figure A.6: Basic cells of device NIN-7 in (a) top-view and (b) cross-section. Some important measurements are shown in orange (in microns). (b) Cross-section of the device along the line indicated by the red arrow in (a). The tightly packed, alternating strips of diffusion regions of opposing polarity are visible. The contacts to these regions do not lie along the line of cross-section and are therefore now shown. The guard ring has also been omitted from this cross-section.

Overall, NIN-7 consists of 2376 alternating strips of p+ and n+ type regions sitting within 12 37 $\times$ 479.6 μm N-well regions. The NW/sub junction totals an area of approximately 184871 μm^2 enclosed by a 11510.4μm long perimeter while the n+/p+ junctions feature a total p+ area of approx. 82970 μm^2 and perimeter of 138283 μm. Interestingly the p+ regions form junctions with their n+ neigbours as well as with the NW basin within which they reside. A rough calculation indicates that a perimeter of approximately 6643.2 μm length is created between p+ and NW regions. A layout view of the basic cell can be found in A.6 with key information summarised in A.7.

NIN-8

NIN-8 (Ninja design family, device identifier no.8) is a device very similar to NIN-7, featuring thin strips of butting diffusion junctions but with the difference that instead of the entire ensemble residing within an N-well region the butting diffusion junctions are implemented directly on substrate. This implies that the p+ regions of the junction act as substrate anchors thus ensuring good ohmic contact to the substrate, much in a similar fashion to n+ regions in NIN-7 which serve as N-well anchors. This design could potentially uncover significant differences in the function of butting junctions on N-well vs similar junctions implemented directly on the substrate. NIN-8 consists of 180 1.2 $\times$ 479.1 μm strips of n+ diffusion that create pn-junctions with interdigitated p+ regions between them and the substrate around and beneath them. This yields an overall junction area of 43356.6 mum^2 and perimeter of 172656 μm. A basic cell as seen in layout view is illustrated in figure A.7 while important design information is

Table A.7: Basic features of device NIN-7

NIN-7 properties table	
Footprint	$479.6 \times 479.6\,\mu m$
Basic cell area (NW)	$15405.94\,\mu m^2$
Basic cell perimeter (NW)	$959.2\,\mu m$
No. of basic cells (NW)	12
Total design area (NW)	$184871.28\,\mu m^2$
Total design perimeter (NW)	$11510.4\,\mu m$
Basic cell area (p+)	$34.92\,\mu m^2$
Basic cell perimeter (p+)	$58.2\,\mu m$
No. of basic cells (p+)	2376
Total design area (p+)	$82969.92\,\mu m^2$
Total design perimeter (p+)	$138283.2\,\mu m$

displayed in table A.8.

Table A.8: Basic features of device NIN-8

NIN-8 properties table	
Footprint	$479.1 \times 479.1\,\mu m$
Basic cell area	$240.92\,\mu m^2$
Basic cell perimeter	$959.2\,\mu m$
No. of basic cells	180
Total design area	$43365.6\,\mu m^2$
Total design perimeter	$172656\,\mu m$

NIN-9

NIN-9 (Ninja design family, device identifier no.9) is a device that was designed with the intention of testing the effects of adding polysilicon above the active area of the device. Nominally, the junction should consist of 60 $5.5 \times 478.3\,\mu m$ strips of n+/sub type sub-junctions, but because AMS35 is a self-aligned process the polysilicon will mask the die while the diffusion process is taking place and therefore shall create diffusion junctions only within those parts of the nominal junction area that are not covered by polysilicon. Ideally this would create basic cells with junctions whose borders are clearly defined either by the edge of the designed diffusion area or by the edges of the polysilicon shapes laid out over the designed diffusion area. This creates a junction with $78766.2\,\mu m$ area and $304134\,\mu m$ perimeter. In reality because of annealing stages throughout the manufacturing process the diffusion regions will invade the space beneath the polysilicon thus distorting the above metrics. Part of a basic cell can be seen in Figure A.8 while important design information is shown in table A.9.

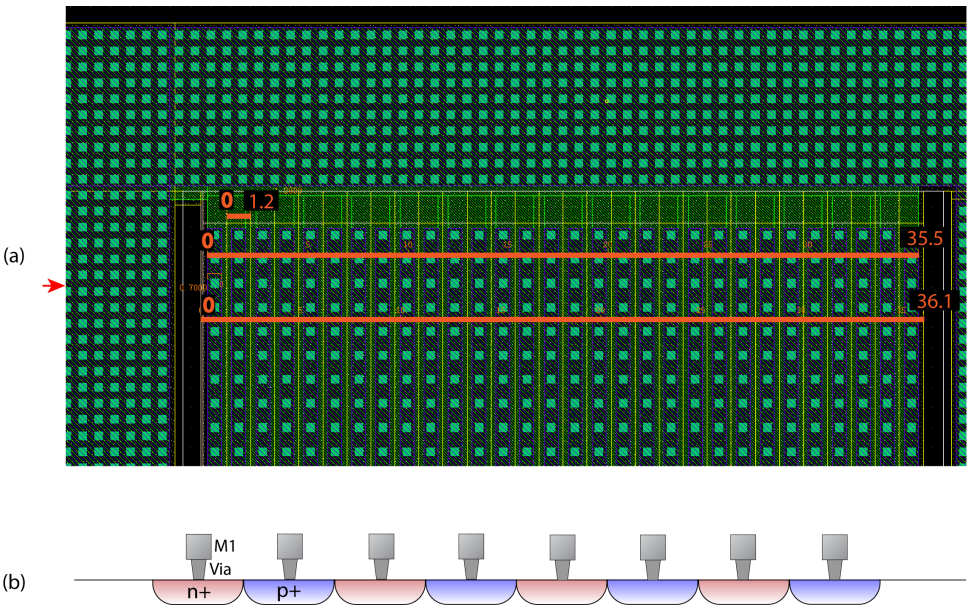

Figure A.7: Basic cell of device NIN-8 in (a) top-view and (b) cross-section. (a) Some important measurements are shown in orange (units of microns). The difference between the 35.5 micron and the 36.1 micron lines lies in that the short line is the width of the nominal diffusion regions and the long line is the width of the diffusion region including outdiffusion. (b) Cross-section of the device along the line indicated by the red arrow in (a).

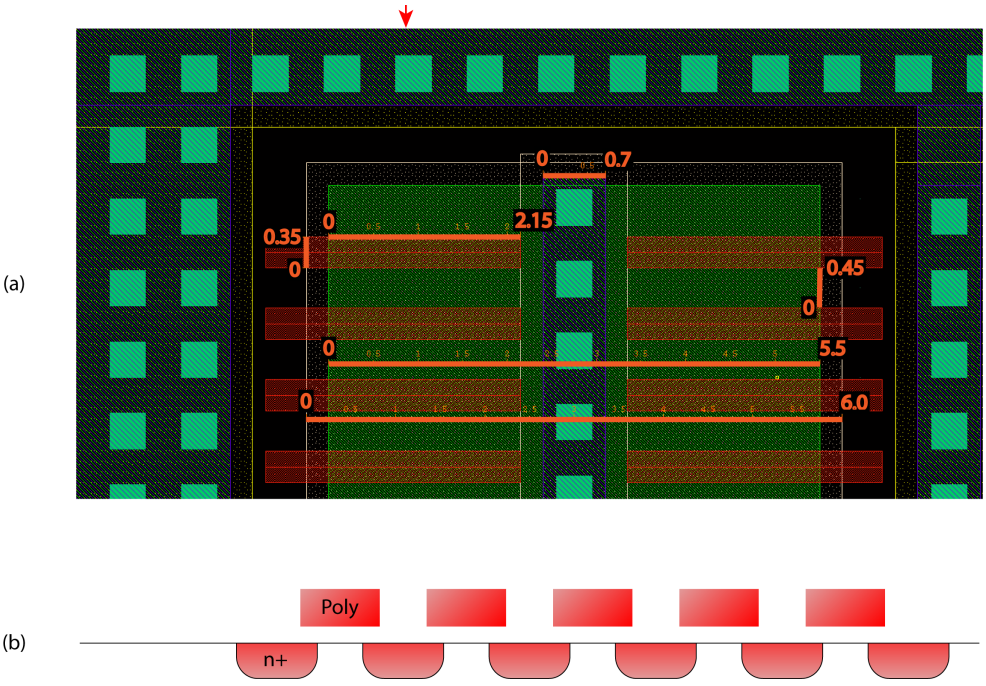

Figure A.8: Basic cell of device NIN-9 in (a) top-view and (b) cross-section. (a) The guard ring can be seen around the core of the device. It consists of p+ diffusion and its corresponding outdiffusion. Strips of polyisilicon used to block the diffusion operation underneath them are visible. (b) Cross-section of the device along the line indicated by the red arrow in (a). The guard ring has been omitted from this panel for clarity. n+ regions are formed in the substrate but are blocked partially by the polysilicon strips lying above them.

Table A.9: Basic features of device NIN-9

NIN-9 properties table	
Footprint	$479 \times 479\,\mu m$
Basic cell area	$1912.77\,\mu m^2$
Basic cell perimeter	$5068.9\,\mu m$
No. of basic cells	80
Total design area	$78766.2\,\mu m^2$
Total design perimeter	$304134\,\mu m$

A.2 The 'Svejk' design family: modulators designed in 0.18 micron technology

The $2.8 \times 3.5 \times 0.25\,mm$ 'Svejk' devices were implemented in IBM's IBM18 0.18 micron technology. This technology offered the ubiquitous N-well and p+/n+ diffusion regions as well as various other 'flavours' of wells, notably the triple, counterdoped p-type well that can be nested in the traditional N-well. The triple well will often be referred to as '3W' for brevity. The Svejk design was tailored with the target of uncovering more subtle differences between devices based on geometry. Thus, devices can be grouped into functional units with specific aims, such as understanding the effects of the side-wall region of a pn-junction by designing a pair of identical pn-junctions and shading the side-wall region selectively in one of them. Also, the contact structures between such 'paired' or otherwise grouped devices have been kept as similar as possible in order to minimise result variation due to contact placement differences.

In total, 12 devices reside on Svejk and it is simple NW/sub, n+/sub and p+/NW/sub devices that form 11 out of the 12 devices. The last one is a 3W/NW/sub device whose role was to uncover potentially interesting quantitative differences between p+/NW and 3W/NW junctions. Exactly 1/2 of the devices (6/12) feature a $300 \times 300\,\mu m$ footprint with the remaining half being limited to a $200 \times 200\,\mu m$ footprint. The choice of sizes and device number was limited by stringent whole-die floorplanning restrictions. A labeled microphotograph of a Svejk specimen can be seen in Figure A.9 whilst the basic properties of the design can be found in table A.10.

Table A.10: Summary of the basic features of the Svejk die family.

Svejk summary	
Technology	IBM18
Feature size	$180\,nm$
Size	$2.8 \times 3.5\,mm$
Die thickness	$250\,\mu m$
Modulator device number	12
Devices sizes	$300 \times 300\,\mu m$ $200 \times 200\,\mu m$

The devices residing upon Svejk are the following (sorted by identifier number):

- SVJ-1: n+/sub device designed to test areal junction response. The entire perimeter of the device is shaded by a metal mask.

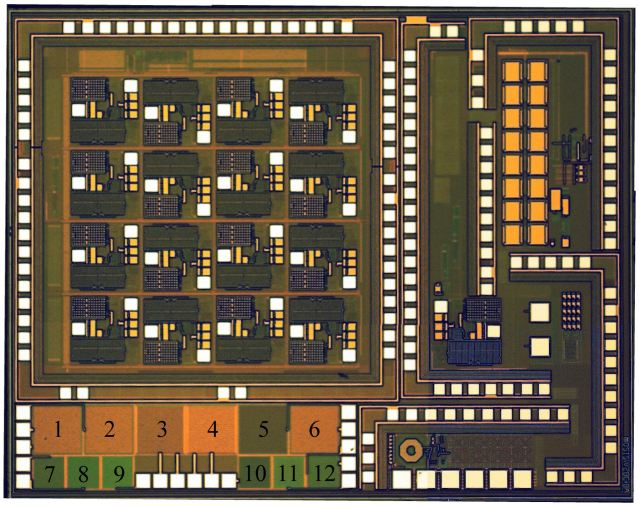

Figure A.9: Svejk device microphotograph with all optical modulator devices labeled by their identification numbers. This was a die shared between more projects. The modulator region seen in the lower left corner measured roughly $2 \times 0.5\,mm$.

- SVJ-2: n+/sub device designed to test areal junction response, similar to SVJ-1. The same metal mask is used for both SVJ-1 and SVJ-2 but in the case of SVJ-2 the diffusion region is allowed to extend for more underneath the mask. The net result is that a pairs of devices with the same areal junction extents are created, but in SVJ-2 the border regions are farther away from exposed area than in SVJ-2 (see corresponding subsections for illustration).
- SVJ-3: Nested p+/NW/sub device where the N-well's entire perimeter is shaded (and therefore acts as a pure areal junction) whilst the p+/NW junction has both areal and side-wall regions exposed.
- SVJ-4: Device very similar to SVJ-3, but this time the metal mask covers much more of the nested junction assembly. In SVJ-4 neither NW/sub or p+/NW junctions have any of their perimeters exposed to light.
- SVJ-5: n+/sub junction with large exposed area and long exposed perimeter.
- SVJ-6: n+/sub junction with large exposed area but the entire perimeter shaded. Very similar to SVJ-5.
- SVJ-7: NW/sub junction featuring very small N-well subunits ($1.5 \times 1.5\,\mu m$ squares).
- SVJ-8: NW/sub junction consisting of small rectangular N-well subunits ($1.5 \times 4.3\,\mu m$ squares).
- SVJ-9: NW/sub junction showcasing medium-size square N-well subunits ($4.3 \times 4.3\,\mu m$ squares).
- SVJ10: NW/sub junction exactly the same as SVJ-8, but with the rectangles oriented in a direction normal to those in SVJ-8.
- SVJ-11: NW/sub junction sporting large square N-well subunits ($9.9 \times 9.9\,\mu m$ squares).
- SVJ-12: Nested 3W/NW/sub junction. The only of its kind on the device.

Therefore the following groups of devices can be discriminated:

1. G1 = SVJ-1, SVJ-2: Pair that can help understand how the presence of a nearby side-wall region can affect areal junctions. Studying n+/sub type junctions.
2. G2 = SVJ-3, SVJ-4: Pair that could reveal the effects of the side-wall area in comparison to those of the areal junction in nested p+/NW/sub junctions.
3. G3 = SVJ-5, SVJ-6: Pair similar to G2 but for simple n+/sub junctions.
4. G4 = SVJ-7, SVJ-8, SVJ-9, SVJ-10, SVJ-11: Quintuplet of NW/sub junctions intended to reveal any aberrations from a simple output vs. geometry relation of the form $O(P, A) = \alpha A + \beta P$ where O is the output, P the perimeter, A the area, and α, β constant coefficients.
5. G5 = SVJ-12: The lone 3W/NW/sub device, intended to reveal any significant modulation efficiency discrepancies with the p+/NW/sub counterparts.

The nomenclature for all devices within a given die is the same and can be seen in table A.11. Please refer to this table as a legend for all layout and cross-section view images of Svejk devices.

Table A.11: Legend to all layout and cross-section view images of devices residing upon the svejk die.

SVEJK IMAGE LEGEND		
	Item	**Explanation**
Layout view	Red (webbed)	Metal layer 1 (M1)
	Yellow (webbed)	Metal layer 2 (M2)
	Green (dotted)	Metal layer 3 (M3)
	Green square	M1 to front end contact
	Pink square	M1 to M2 contact
	Purple square	M2 to M3 contact
	Cyan (dotted)	Diffusion region
	White (dotted)	Turns diffusions to p+ type
	Purple (dotted)	N-well region
	Light green (dotted)	3W region
Cross-section view	Mx	X-th metallisation layer
	Via	Connection between metallisation layers or metallisation and substrate
	n+/p+	n+/p+ diffusion regions
	NW	N-well
	3W	Triple well (p-type)

SVJ-1

SVJ-1 (Svejk design family, device identifier no.1) is an n+/sub device that consists of 900 relatively small 4×3.96 micron rectangular n+ regions sitting atop the substrate. This is the size of the exposed area within each basic cell only; the full extent of the diffusion region belonging to each basic cell is a 6×6 microns square instead. SVJ-1 is intended for comparison with SVJ-2 and its specific purpose is to uncover the behaviour of the areal component of the n+/sub junction when the entire area is relatively close to the side-wall components. In total approximately $11615\,\mu m^2$ of area are exposed to illumination in this design. No part of the perimeter is exposed, however. A layout and a cross-section view of the basic cell of SVJ-1 can be seen in Figure A.10 with key statistics summarised in table A.12.

Table A.12: Basic features of device SVJ-1.

SVJ-1 properties table	
Group membership	G1
Footprint	$300 \times 300\,\mu m$
Basic cell area	$12.9056\,\mu m^2$
Basic cell perimeter	$0\,\mu m$
No. of basic cells	900
Total design area	$11615.04\,\mu m^2$
Total design perimeter	$0\,\mu m$

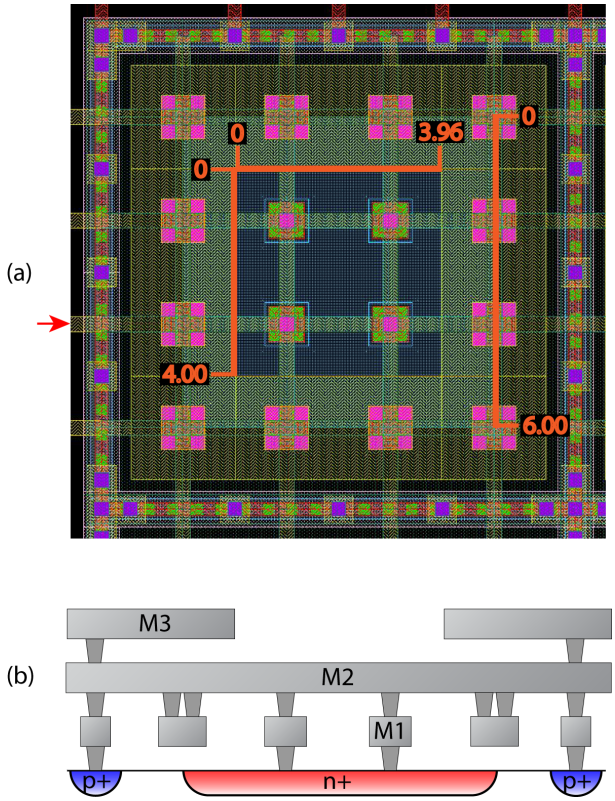

Figure A.10: Basic cell of device SVJ-1 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show the exposed area (3.96×4.00 microns) and the underlying diffusion region extent (6 microns square). The guard ring is also visible. (b) Visible features in this cross-sections include the M3 mask, the guard ring, the supply line that provides the biasing (on M2) and the extent of the diffusion region.

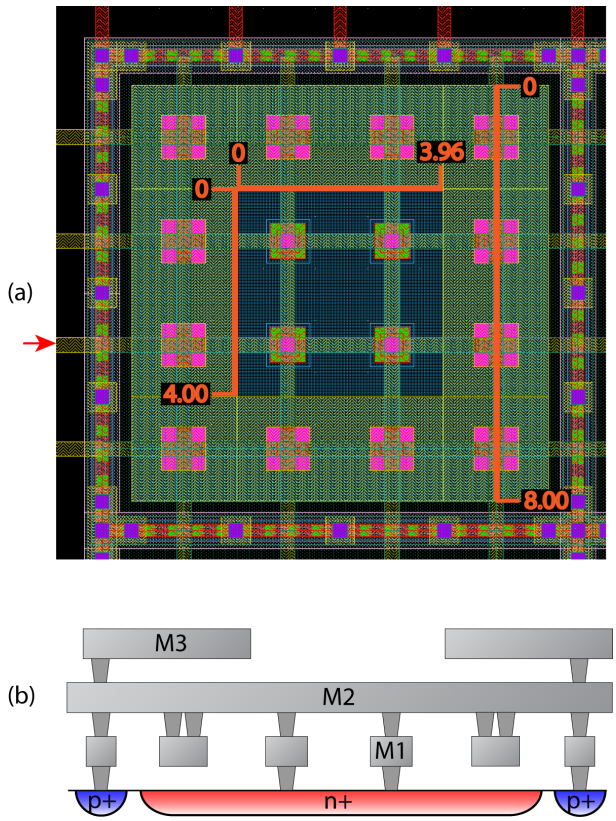

Figure A.11: Basic cell of device SVJ-2 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show the exposed area (3.96×4.00 microns) and the underlying diffusion region extent (8 microns square). The guard ring is also visible. (b) Visible features in this cross-sections include the M3 mask, the guard ring, the supply line that provides the biasing (on M2) and the extent of the diffusion region.

SVJ-2

SVJ-2 (Svejk design damily, device identifier no.2) is an n+/sub device that is identical to SVJ-1 in every respect but the extent of the diffusion region beneath the metal mask that determines the exposed area of the device. As such, it too consists of 900 4×3.96 micron n+ type basic cells with a total exposed area of 11615 μm^2 even though each basic cell now features diffusion regions extending for 8×8 microns. SVJ-2 can facilitate the understanding of the behaviour of n+/sub junction areal components when the exposed areal component is relatively far away from the side-wall component. To this end comparing SVJ-2 performance with SVJ-1 can prove educational. A layout and a cross-section view of the basic cell of SVJ-2 can be seen in Figure A.11 with key statistics summarised in table A.13.

SVJ-3

SVJ-3 (Svejk design family, device identifier no.3) in a nested p+/NW/sub junction with the characteristic feature that the p+/NW sub-junction is almost fully exposed to the incoming light (except for areas shaded by the contacts that link the device to the bond pads of the die) whilst the NW/sub junction side-walls are shaded from light but most of the areal component is exposed. Thus, SVJ-3

Table A.13: Basic features of device SVJ-2.

SVJ-2 properties table	
Group membership	G1
Footprint	$300 \times 300\,\mu m$
Basic cell area	$12.9056\,\mu m^2$
Basic cell perimeter	$0\,\mu m$
No. of basic cells	900
Total design area	$11615.04\,\mu m^2$
Total design perimeter	$0\,\mu m$

consists of 576 basic cells, each of which consists of a single N-well and a single p+ diffusion region. The p+/NW component features a total exposed area of approx. $30627.5\,\mu m^2$ and a total exposed perimeter of $16865.3\,\mu$ while the NW/sub component features an exposed area measuring $31819.4\,\mu m^2$ with no exposed perimeter. This configuration allows for the simultaneous study of the behaviour of a full (areal and side-wall components) p+/NW junction and the areal component of the NW/sub junction without much interference from the NW side-walls. Layout and cross-section views of SVJ-3 can be seen in Figure A.12 with basic statistics of the device summarised in table A.14.

Table A.14: Basic features of device SVJ-3.

SVJ-3 properties table	
Group membership	G2
Footprint	$300 \times 300\,\mu m$
Basic cell area (NW)	$55.242\,\mu m^2$
Basic cell perimeter (NW)	$0\,\mu m$
No. of basic cells (NW)	900
Total design area (NW)	$31819.392\,\mu m^2$
Total design perimeter (NW)	$0\,\mu m$
Basic cell area (p+)	$53.1728\,\mu m^2$
Basic cell perimeter (p+)	$29.28\,\mu m$
No. of basic cells (p+)	900
Total design area (p+)	$30627.5328\,\mu m^2$
Total design perimeter (p+)	$16865.28\,\mu m$

SVJ-4

SVJ-4 (Svejk design family, device dientifier no.4) is a p+/NW/sub nested junction device with its main characteristic being that neither NW/sub nor p+/NW components have any of their perimeter exposed to incoming light. This is achieved by use of an extensive mask. Aside from this difference in masking, SVJ-4 is identical to SVJ-3 in design, i.e. their front ends are indistinguashable. Thus, SVJ-4 consists of 576 basic cells where by both NW/sub and p+/NW junctions feature an exposed area of $6757.1712\,\mu m^2$ and an exposed perimeter of $0 \mu m$. This design, when examined in tandem with SVJ-3 should be theoretically capable of revealing the influence of side-wall junctions of the p+/NW type on overall modulator performance, at least for the given separation between the edges of the p+ region and the edges of the N-well. A layout and a cross-section view can be seen in Figure A.13 with important metrics of SVJ-4 summarised in table A.15.

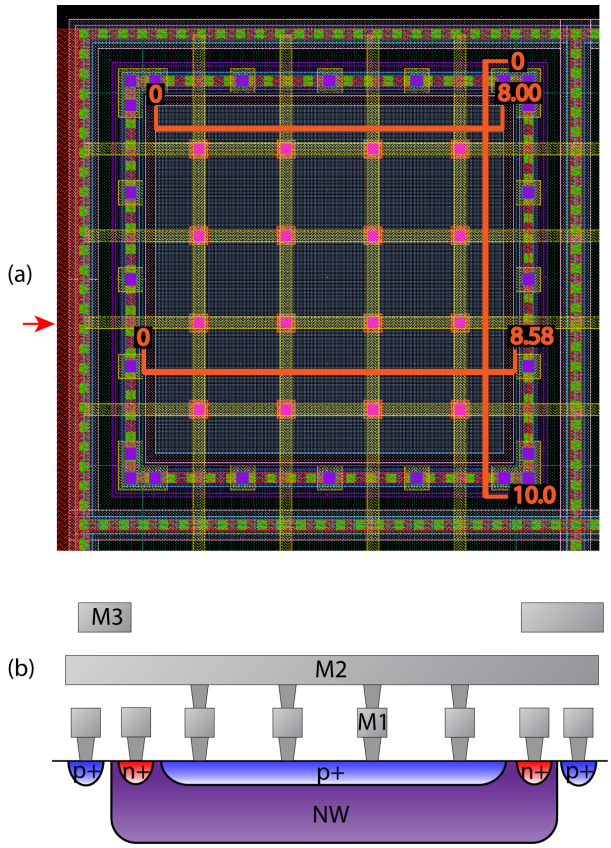

Figure A.12: Basic cells of device SVJ-3 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show the total area of both the p+/NW subjunction (8 microns square) and for NW/sub subjunction (10 microns square) and the exposed area extent (8.58 microns square). The guard ring is also visible. (b) Visible features in this cross-sections include the M3 mask, the guard ring, the supply line that provides the biasing to the diffusion region (on M2), the extent of the diffusion region, the extent of the well, and the contacts that link the N-well to its corresponding bond pad.

Table A.15: Basic features of device SVJ-4.

SVJ-4 properties table	
Group membership	G2
Footprint	$300 \times 300\,\mu m$
Basic cell area (NW)	$11.7312\,\mu m^2$
Basic cell perimeter (NW)	$0\,\mu m$
No. of basic cells (NW)	900
Total design area (NW)	$6757.1712\,\mu m^2$
Total design perimeter (NW)	$0\,\mu m$
Basic cell area (p+)	$11.7312\,\mu m^2$
Basic cell perimeter (p+)	$0\,\mu m$
No. of basic cells (p+)	900
Total design area (p+)	$6757.1712\,\mu m^2$
Total design perimeter (p+)	$0\,\mu m$

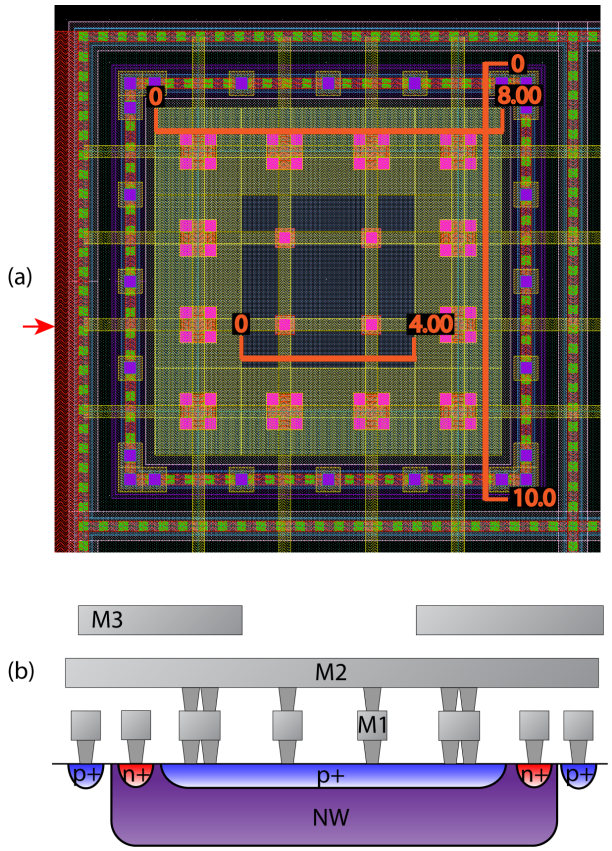

Figure A.13: Basic cells of device SVJ-4 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show the total area of both the p+/NW subjunction (8 microns square) and for NW/sub subjunction (10 microns square) and the exposed area extent (8.58 microns square). The guard ring is also visible. (b) Visible features in this cross-sections include the M3 mask, the guard ring, the supply line that provides the biasing to the diffusion region (on M2), the extent of the diffusion region, the extent of the well, and the contacts that link the N-well to its corresponding bond pad. Note: as can be seen from the layout view, M2 also plays a role as a masking layer, however due to the choice of cross-section location this is not visible in the cross-section view.

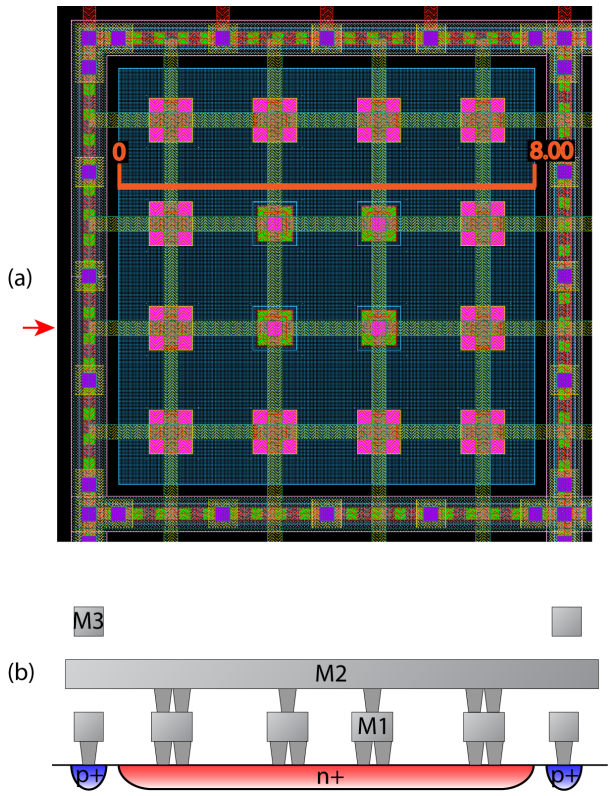

Figure A.14: Basic cell of device SVJ-5 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bar shows the extent of the n+/sub junction (8 microns square). The guard ring is also visible. (b) Notable features in this cross-sections include the guard ring, the supply line that provides the biasing to the diffusion region (on M2) and the extent of the diffusion region.

SVJ-5

SVJ-5 (Svejk design family, device identifier no.5) is an n+/sub type device whereby the basic cell consists of a fully exposed junction. Aside from the necessary metallisation that services the electrical contacts to the device no masking of any sort was employed in this device. Thus, a design based on 900 8×8 micron basic cells was created with a total exposed area of approx. 38698 μm^2 and perimeter of 24768 μm. Interestingly, despite the absence of a mask the exposed area of each basic cell measures only about 42.9 μm^2 out of the nominal 64 μm^2 which confirms that contact metallisation cannot be neglected from the calculations. SVJ-5 is suitable for uncovering the effects of side-wall regions of the n+/sub junction when compared to SVJ-6 (see next subsection). Layout and cross-section views are shown in Figure A.14 with key information summarised in A.16

SVJ-6

SVJ-6 (Svejk design family, device identifier no.6) is an n+/sub device that is paired to SVJ-5. As such the pair shares a front end design, but differs in that SVJ-6 possesses a mask that shades the perimeter of the junction within each basic cell. This implies that the exposed area is smaller compared to SVJ-5 and the exposed perimeter length is 0. The nominal size of the diffusion region in each basic cell measures

Table A.16: Basic features of device SVJ-5.

SVJ-5 properties table	
Group membership	G3
Footprint	$300 \times 300\,\mu m$
Basic cell area	$42.9312\,\mu m^2$
Basic cell perimeter	$27.52\,\mu m$
No. of basic cells	900
Total design area	$38638.08\,\mu m^2$
Total design perimeter	$24768\,\mu m$

$8 \times 8\,\mu m$ and contrasts with the $5.72 \times 5.72\,\mu m$ exposed window. In total, the 900 basic cells of SVJ-6 feature an exposed area of $21975.84\,\mu m^2$. Having the side-wall regions of the diffusion junction masked, SVJ-6 can help understand the influence of the side-wall segment of the junction vs. the contribution from its areal component, if SVJ-6 modulator performance is compared to that of the SVJ-5 device. Figure A.15 shows a layout and a cross-section view of the device. Important data on SVJ-6 can be seen in table A.17.

Table A.17: Basic features of device SVJ-6.

SVJ-6 properties table	
Group membership	G3
Footprint	$300 \times 300\,\mu m$
Basic cell area	$24.4176\,\mu m^2$
Basic cell perimeter	$0\,\mu m$
No. of basic cells	900
Total design area	$21975.84\,\mu m^2$
Total design perimeter	$0\,\mu m$

SVJ-7

SVJ-7 (Svejk design family, device identifier no.7) is an NW/sub type device featuring the smaller $200 \times 200\,\mu m$ footprint as opposed to the larger $300 \times 300\,\mu m$ devices SVJ-1 to SVJ-6. SVJ-7 is part of a series of devices intended for understanding the finer behaviour of N-well structures and its specific role is to determine the electro-optical characteristics of the minimum size well, which in the IBM18 technology measures 1.5 microns across. This accounts for the small size of the basic cell. In total 4096 basic cells combine to create a total exposed area of $7045.12\,\mu m^2$ with total exposed perimeter of $22282.24\,\mu m$. A layout and a cross-section view of the device can be seen in Figure A.16 with key information summarised in table A.18.

SVJ-8

SVJ-8 (Svejk design family, device identifier no.8) is an NW/sub device featuring a rectangular basic cell. The 1.5 micron width of the cell is identical to the side length of the basic cell of SVJ-7, but the height measures $4.3\,\mu m$. The design pattern of the SVJ-8 basic cell can be conceptually thought of as the

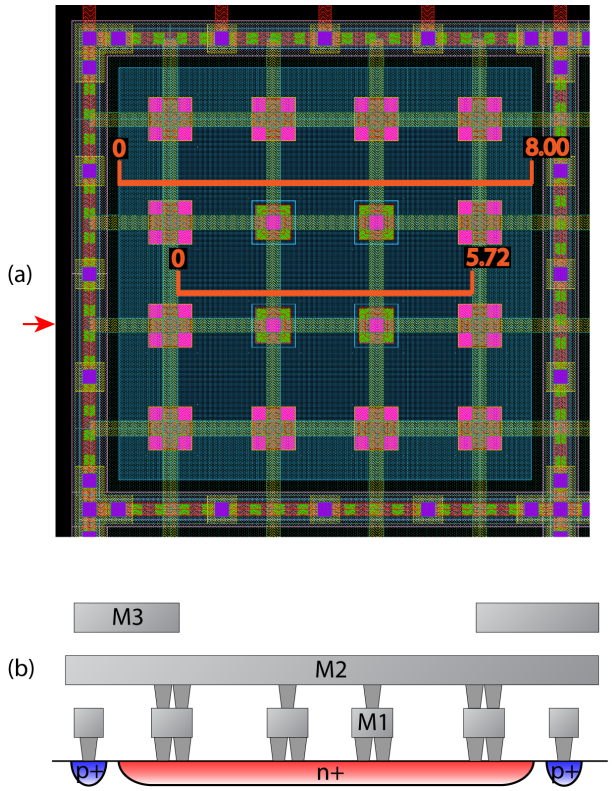

Figure A.15: Basic cell of device SVJ-6 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bar shows the extent of the n+/sub junction (8 microns square). The guard ring is also visible. (b) Notable features in this cross-sections include the M3 mask, the guard ring, the supply line that provides the biasing to the diffusion region (on M2) and the extent of the diffusion region.

Table A.18: Basic features of device SVJ-7.

SVJ-7 properties table	
Group membership	G4
Footprint	$200 \times 200\,\mu m$
Basic cell area	$1.72\,\mu m^2$
Basic cell perimeter	$5.44\,\mu m$
No. of basic cells	4096
Total design area	$7045.12\,\mu m^2$
Total design perimeter	$22282.24\,\mu m$

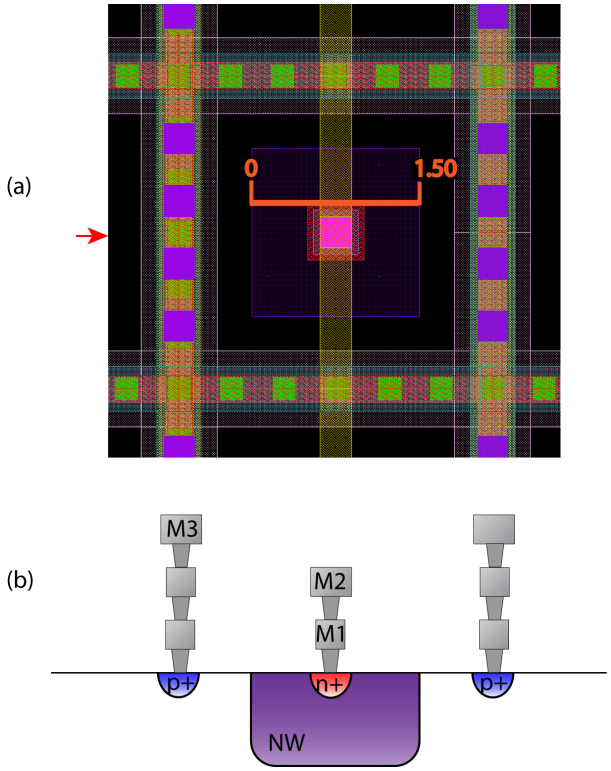

Figure A.16: Basic cell of device SVJ-7 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bar shows the extent of the NW/sub junction. The guard ring is also visible. (b) Notable features in this cross-sections include the guard ring, the supply line that provides the biasing to the diffusion region (on M2) and the extent of the well region.

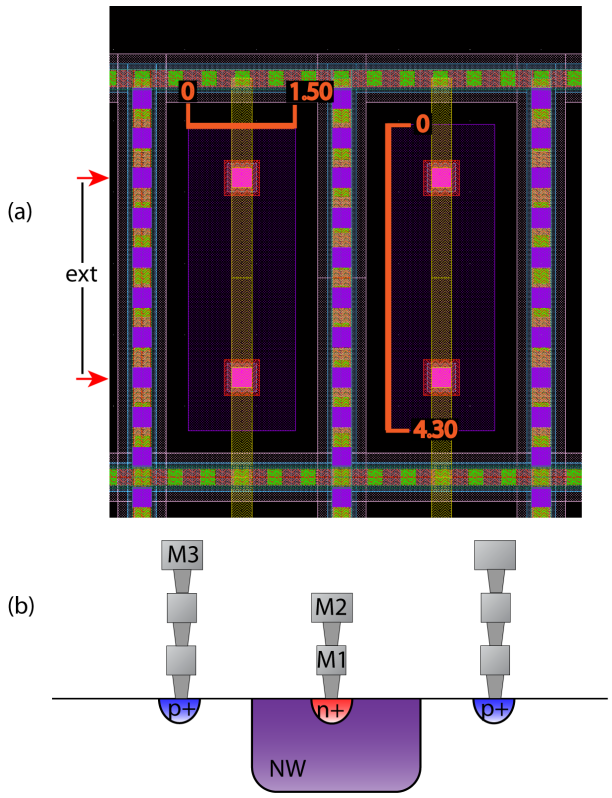

Figure A.17: Basic cell pair of device SVJ-8 in (a) top-view and (b) cross-section along the lines indicated by the red arrows in (a) for only one of the devices in the pair. The cross-sections would be identical along both arrows which also serve to illustrate the limits of the extension that transforms SVJ-7 into SVJ-8 (bar marked 'ext'). (a) The orange bars show the extent of the NW/sub junction. The guard ring is also visible. (b) Notable features in this cross-sections include the guard ring, the supply line that provides the biasing to the diffusion region (on M2) and the extent of the well region.

pattern of SVJ-7 trombone-extended by 2.8 microns across a line cutting horizontally through the middle of the device. The extension consists of an N-well and the halves of the contacts that would be missing, had the extension been realised using only N-well material. This concept is illustrated along with the layout and cross-sectional views in figure A.17. In total, 2112 basic cells offer an exposed area covering $10614.942\,\mu m^2$ with a corresponding exposed perimeter of $23316.48\,\mu m$. SVJ-8 should in theory yield an estimate of how the added 'tube' of N-well affects modulator operation when compared to SVJ-7. The added tube consists of two side-wall and an areal component, but notably no corners. Important information concerning SVJ-8 can be found in table A.19.

SVJ-9

SVJ-9 (Svejk design family, device identifier no.9) is another NW/sub device that belongs to group G4. It features a square basic cell of medium size ($4.3 \times 4.3\,\mu m$) and can be regarded as a trombone-extended version of SVJ-8 by 2.8 microns along a line cutting SVJ-8 vertically in half. This is very similar to how SVJ-8 was generated as an extension of SVJ-7. Thus, SVJ-8 is theoretically suited for revealing how a finite length 'tube' of N-well (two side-walls and one areal component) can affect modulator performance. This tube is thicker than its SVJ-8 equivalent and therefore its areal component is more prominent. In

Table A.19: Basic features of device SVJ-8.

SVJ-8 properties table	
Group membership	G4
Footprint	$200 \times 200\,\mu m$
Basic cell area	$6.45\,\mu m^2$
Basic cell perimeter	$1.424\,\mu m$
No. of basic cells	2112
Total design area	$10614.942\,\mu m^2$
Total design perimeter	$23316.48\,\mu m$

total, 1024 basic cells enclose an exposed area of approx. $16017.4\,\mu m^2$ with an exposed perimeter of $16465.92\,\mu m$. Layout and cross-section views are shown in Figure A.18. Important data is summarised in table A.20.

Table A.20: Basic features of device SVJ-9.

SVJ-9 properties table	
Group membership	G4
Footprint	$200 \times 200\,\mu m$
Basic cell area	$15.642\,\mu m^2$
Basic cell perimeter	$16.08\,\mu m$
No. of basic cells	1024
Total design area	$16017.408\,\mu m^2$
Total design perimeter	$16465.92\,\mu m$

SVJ-10

SVJ-10 (Svejk design family, device identifier no.10) is an NW/sub device most closely linked to SVJ-10 with which it shares the entire design, save for the direction of the basic cells. In SVJ-8 the long side of the basic cells is oriented in a conventionally called 'vertical' orientation whilst in SVJ-10 they have been turned by a right angle in order to assume a 'horizontal' orientation. This could reveal potential differences or variation in modulator performance arising from the orientation of the basic cells that are added on top of the normal device-to-device variation one would expect to see anyway. Unfortunately, the proximity of the photosensitive area to a corner of the chip could imply that there might be severe stress gradients across much of its surface and thus much of the performance discrepancy between SVJ-8 and SVJ-10 could be reasonably attributed to the chip warping effect. The layout and cross-section views are shown in Figure A.19 while important device stats reside in table A.21.

SVJ-11

SVJ-11 (Svejk design family, device identifier no.11) is the last NW/sub device of the G4 group and features the largest exposed area and the smallest corresponding perimeter of all devices belonging to G4. It can be seen as the 'logical magnification' of SVJ-9 but with the contacts remaining spaced exactly as they did in SVJ-9. Therefore the net difference between SVJ-9 and SVJ-11 is that if the well region of

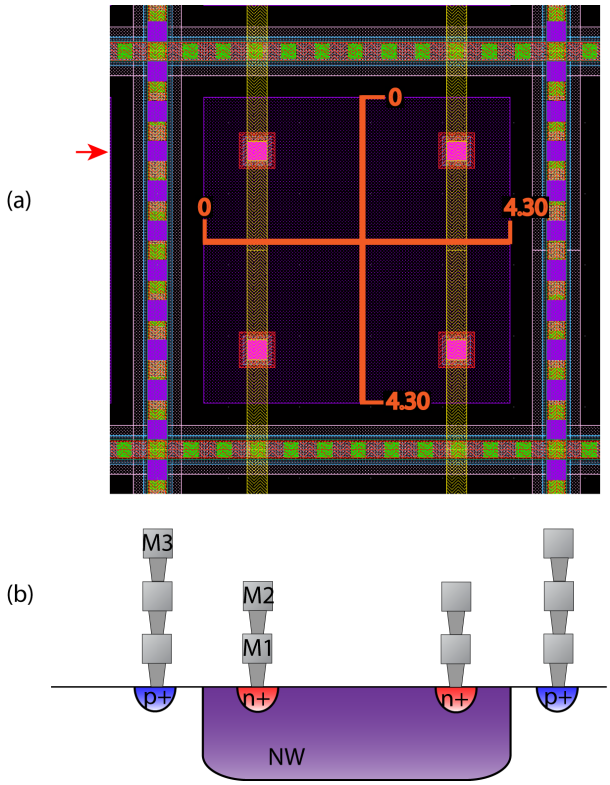

Figure A.18: Basic cell of device SVJ-9 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show the extent of the NW/sub junction. The guard ring is also visible. (b) Notable features in this cross-sections include the guard ring, the two supply lines that provide the biasing to the diffusion region (on M2) and the extent of the well region.

Table A.21: Basic features of device SVJ-10.

SVJ-10 properties table	
Group membership	G4
Footprint	$200 \times 200\,\mu m$
Basic cell area	$6.45\,\mu m^2$
Basic cell perimeter	$1.424\,\mu m$
No. of basic cells	2112
Total design area	$10614.942\,\mu m^2$
Total design perimeter	$23316.48\,\mu m$

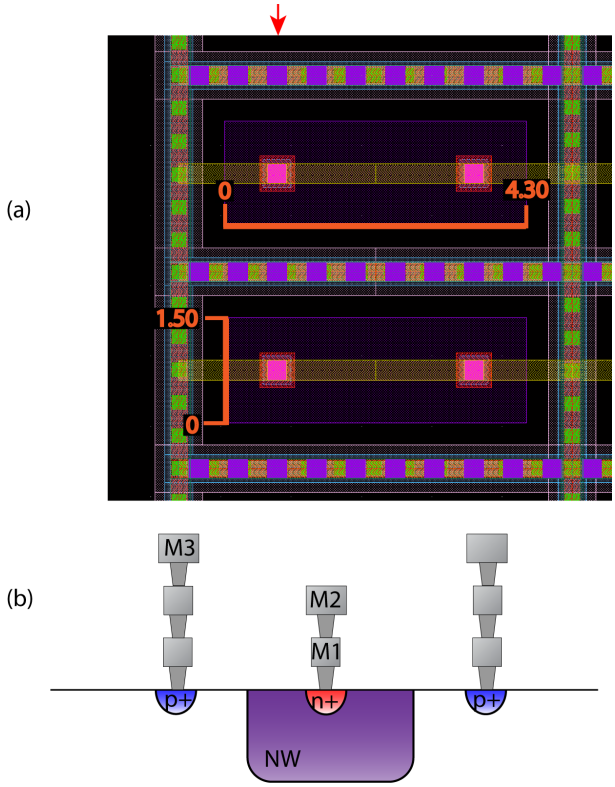

Figure A.19: Basic cell pair of device SVJ-10 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a) for only one of the devices in the pair. (a) The orange bars show the extent of the NW/sub junction. The guard ring is also visible. (b) Notable features in this cross-sections include the guard ring, the supply line that provides the biasing to the diffusion region (on M2) and the extent of the well region.

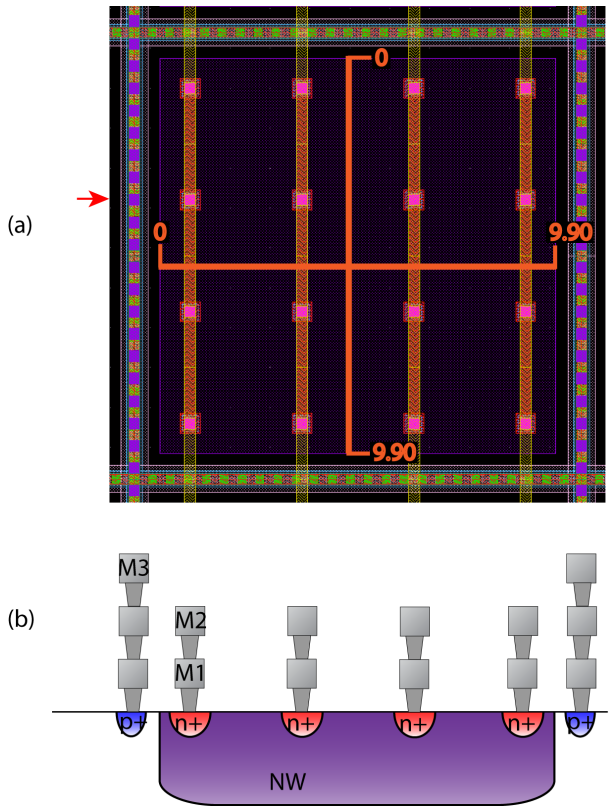

Figure A.20: Basic cell of SVJ-11 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show the extent of the NW/sub junction. The guard ring is also visible. (b) Notable features in this cross-sections include the guard ring, the four supply lines that provide the biasing to the diffusion region (on M2) and the extent of the well region.

SVJ-11 is to be split into Voronoi cells on the basis of the well contacts[2] then SVJ-9 would consist of the V-cells belonging to the four corner contacts of SVJ-11 and the rest would be the extensions. This cross-shaped extension area will be dominated by its areal junction component (at least in theory). A total of 256 basic cells enclose an exposed area of $24524.288\,\mu m^2$ with an exposed perimeter of $9553.92\,\mu m$. Layout and cross-sectional views can be found in Figure A.20 with key information in table A.22.

Table A.22: Basic features of device SVJ-11.

SVJ-11 properties table	
Group membership	G4
Footprint	$200 \times 200\,\mu m$
Basic cell area	$95.798\,\mu m^2$
Basic cell perimeter	$37.32\,\mu m$
No. of basic cells	256
Total design area	$24524.88\,\mu m^2$
Total design perimeter	$9553.92\,\mu m$

[2] A Voronoi cell is a subdivision of a plane (or volume) whereby each point within the cell belongs to its nearest reference point. Reference points may be distributed in any fashion on the plane and each features its own Voronoi cell.

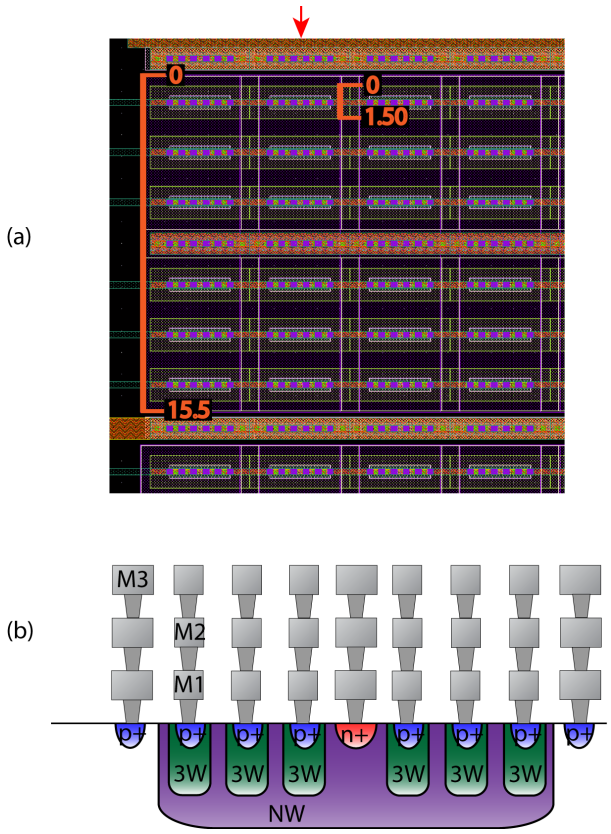

Figure A.21: Part of N-well/sub basic cell of device SVJ-12 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show the width of the NW/sub junction (length = 175.7 μm) and of the 3W/NW junction (length = 174.8μm). The guard ring is also visible. (b) Notable features in this cross-section include the substrate anchors, the N-well anchor (running along the middle of the N-well), the six supply lines that provide the biasing to the 3W regions (on M3) and the extent of the well region.

SVJ-12

SVJ-12 (Svejk design family, device identifier no.12), the last device of the Svejk design is a 3W/NW/sub nested junction device. It belongs to a category of its own and is intended to reveal any potentially large differences in the efficiency of a 3W/NW modulator vs a p+/NW one or perhaps any possibly significant differences between the NW/sub component of the junction introduced by the presence of the 3W. The NW/sub basic cell of this device is also unique amongst other SVJ devices in that it is approx. 175 microns long, i.e. takes the form of a long strip. More specifically it consists of a large N-well that contains exactly six 3W regions within it. In total the NW/sub junction covers a total exposed area of 24559.15 μm^2 with corresponding perimeter of 4175.6 μm whilst the 3W/NW component accounts for 13844.16 μm^2 of area and a 23232 μm perimeter. A layout and cross-section view diagram is provided in Figure A.21 with important data summarised in table A.23.

Table A.23: Basic features of device SVJ-12. For the purposes of this table each 3W region is considered as its own basic cell.

SVJ-12 properties table	
Group membership	G5
Footprint	$200 \times 200\,\mu m$
Basic cell area (NW/sub)	$2232.65\,\mu m^2$
Basic cell perimeter (NW/sub)	$379.6\,\mu m$
No. of basic cells (NW/sub)	11
Total design area (NW/sub)	$24559.15\,\mu m^2$
Total design perimeter (NW/sub)	$4175.6\,\mu m$
Basic cell area (3W/NW)	$209.76\,\mu m^2$
Basic cell perimeter (3W/NW)	$352.0\,\mu m$
No. of basic cells (3W/NW)	66
Total design area (3W/NW)	$13844.16\,\mu m^2$
Total design perimeter (3W/NW)	$23232\,\mu m$

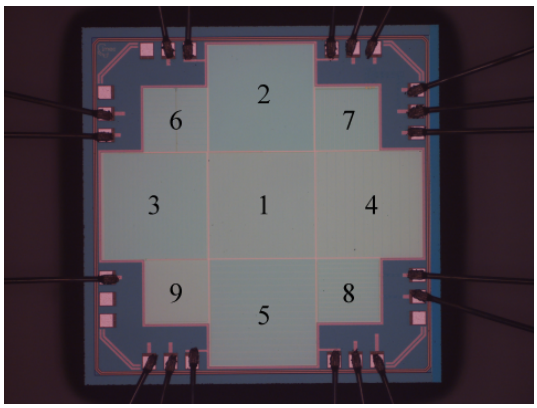

Figure A.22: Teddy device microphotograph with all optical modulator devices labeled by their identification numbers.

A.3 The 'Teddy' design family: modulators designed in 0.13 micron technology

The $1.525 \times 1.525 \times 0.279\,mm$ 'Teddy' devices were implemented in UMC's UMC13 0.13 micron technology. UMC13 offers the standard n+/sub, p+NW and NW/sub junction types, but also allows for manufacturing of butting diffusion n+/p+ junctions (either directly on substrate or within the confines of an N-well) as well as 3W-based devices. The Teddy was designed primarily as an extension to the Ninja design with the aim of uncovering important relations between junction geometry and type vs performance within a 0.13 micron node context. As such, each Teddy die hosts nine devices grouped into four pairs and a singleton device. No metal masking was employed in this design. The devices can either be of the 'large' sort, measuring approx. $500 \times 500\,\mu m$ or of the 'small' sort, measuring $300 \times 300\,\mu m$. The choice of device sizes was such as to maximise useful device area while wasting as little Silicon area as possible. A microphotograph of a sample Teddy die can be seen in Figure A.22. Notably, the devices belonging to the small variety are sized and positioned so that the bond pads can fit comfortably at the corners of the die (where the largest stress gradients are also observed). Key metrics of the Teddy die are summarised in table A.24.

Table A.24: Summary of the basic features of the Teddy die family.

Teddy summary	
Technology	UMC13
Feature size	$130\,nm$
Size	$1.525 \times 1.525\,mm$
Die thickness	$279\,\mu m$
Modulator device number	9
Devices sizes	$500 \times 500\,\mu m$ $300 \times 300\,\mu m$

The devices residing upon Svejk are the following (sorted by identifier number):

- TED-1: An NW/sub device featuring a small basic cell (by area).
- TED-2: An n+/sub device featuring a small basic cell.
- TED-3: An n+/sub device featuring a large basic cell.
- TED-4: An NW/sub device featuring a large basic cell.
- TED-5: A nested 3W/NW/sub device featuring small 3W/NW basic cells.
- TED-6: The only butting diffusion n+/p+ device, laid out directly on substrate.
- TED-7: Nested n+/3W/NW/sub device featuring large 3W/NW and n+/3W basic cells.
- TED-8: Nested n+/3W/NW/sub device featuring small 3W/NW and n+/3W basic cells.
- TED-9: Nested 3W/NW/sub device featuring large 3W/NW basic cells.

The size of the basic cells can be described either as large or small simply because such differentiation will affect their areal/side-wall component ratios. Thus when two devices consisting of the same junction types but designed with different area/side-wall component ratios are paired, i.e. their performances are measured and compared to each other, then some clues as to the specific influence of areal and side-wall junctions should become evident. Therefore the devices described above can be naturally grouped in the following categories:

- H1 = TED-1, TED-4: NW/sub junction pair.
- H2 = TED-2, TED-3: n+/sub junction pair.
- H3 = TED-5, TED-9: TW/NW/sub junction pair.
- H4 = TED-7, TED-8: n+/3W/NW/sub junction pair.
- H5 = TED-6: The butting diffusion device alone.

The nomenclature for all devices within a given die is the same and can be seen in table A.25. Please refer to this table as a legend for all layout and cross-section view images of Teddy devices.

Table A.25: Legend to all layout and cross-section view images of devices residing upon the Teddy die.

TEDDY IMAGE LEGEND		
	Item	**Explanation**
Layout view	Cyan (striped)	Metal layer 1 (M1)
	Yellow (striped)	Metal layer 2 (M2)
	Green square	M1 to front end contact
	Red (solid)	Diffusion region
	Orange box (outline)	Turns diffusions to n+ type
	Pink box (outline)	Turns diffusions to p+ type
	White box (outline)	N-well region
	Green box (outline)	3W region
Cross-section view	Mx	X-th metallisation layer
	Via	Connection between metallisation layers or metallisation and substrate
	n+/p+	n+/p+ diffusion regions
	NW	N-well
	3W	Triple well (p-type)

TED-1

TED-1 (Teddy design family, device identifier no.1) is an NW/sub junction that consists of 30788 basic cells measuring $1.24 \times 1.24\,\mu m$. The small size of the basic cells gives rise to the vast combined junction perimeter measuring $133004.16\,\mu m$ with a relatively small junction area of $35122.9504\,\mu m^2$. TED-1 weas designed as the counterpart of TED-4, a device featuring the same junction type but with different area/perimeter ratio. By comparing their performances, the individual contributions of the areal and side-wall components should become evident. A combined layout and cross-section view of TED-1 is shown in Figure A.23 with basic information summarised in table A.26.

Table A.26: Basic features of device TED-1.

TED-1 properties table	
Group membership	H1
Footprint	$500 \times 500\,\mu m$
Basic cell area	$1.1408\,\mu m^2$
Basic cell perimeter	$4.32\,\mu m$
No. of basic cells	30788
Total design area	$35122.9504\,\mu m^2$
Total design perimeter	$133004.16\,\mu m$

TED-2

TED-2 (Teddy design family, device identifier no.2) is an n+/sub device designed on the basis of a very small basic cell. In total 96300 $1.32 \times 1.32\,\mu m$ basic cells combine to yield an area of $127116\,\mu m^2$ with corresponding perimeter of $446832\,\mu m$. The closer minimum allowable spacing between diffusion regions allows the density of basic cells on TED-2 to greatly exceed its corresponding value in TED-1. This accounts for the substantially larger area and perimeter that TED-1 exhibits. Following the minimum

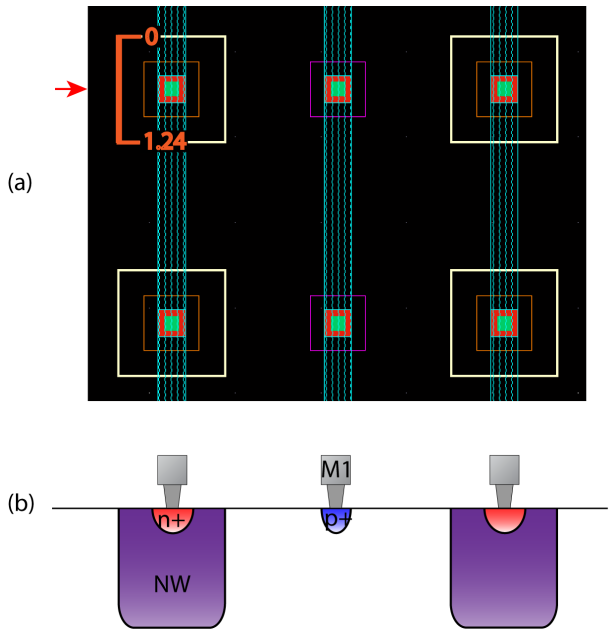

Figure A.23: Basic cell of device TED-1 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bar shows the extent of the NW/sub junction. Also visible are a couple of substrate anchor contacts that keep the substrate biased at GND in their neighbourhood. (b) Notable features in this cross-section include the substrate anchor and the extent of the well regions.

spacing rule should still guarantee that each basic cell does not interact with neighbouring basic cells in order to produce a single 'super-cell', that is to say between any pair of neighbouring basic cells there should be a strip of net p-type doping as opposed to a continuous extent of net n-type material. TED-2's counterpart is TED-3, which features the same junction type but a dramatically reduced perimeter and an enlarged areal component. Together, TED-2 and TED-3 should reveal the individual contributions of side-wall and areal components of n+/sub junctions. Layout and cross-section views can be found in Figure A.24 while important data is summarised in table A.27.

Table A.27: Basic features of device TED-2.

TED-2 properties table	
Group membership	H2
Footprint	$500 \times 500\,\mu m$
Basic cell area	$1.32\,\mu m^2$
Basic cell perimeter	$4.64\,\mu m$
No. of basic cells	96300
Total design area	$127116\,\mu m^2$
Total design perimeter	$446832\,\mu m$

TED-3

TED-3 (Teddy design family, device dientifier no.3) is an n+/sub junction paired with TED-2. It's main characteristic is the large size of its diffusion regions that measure $20 \times 491\,\mu m$ in size. This gives rise to an areal component measuring $199620.96\,\mu m^2$ and a perimeter that stands at a comparatively small

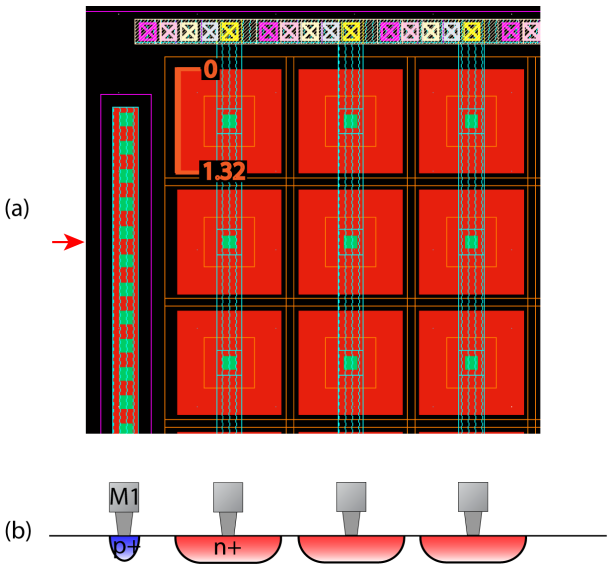

Figure A.24: Basic cell of device TED-2 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bar shows the extent of the n+/sub junction. Also visible is a substrate anchor contact that keeps the substrate biased at GND in its neighbourhood (far left) and a bundle of metallisation lines spanning all available metal layers (top side of the image). The latter structure provides the contact between the device and its corresponding bond pad and is overdesigned in such manner to ensure that large currents can flow in and out of the gigantic (by CMOS standards) device without electromigration and track melting issues. (b) Notable features in this cross-section include the substrate anchor and the extent of the well regions.

21435.12 μm. Each such diffusion region is a basic cell and 21 of these cells cover the entire surface of the device. The 20 μm width of the basic cells was chosen to have that specific value in order to allow for sufficient substrate anchor contacts and thus maintain the substrate under and around the basic cells at a potential that can be regarded as reasonably close to GND. Thus substrate contacts flank basic cells all along their long sides. Contact supply lines to the diffusion region itself were kept close to the edges so as not to interfere with the doping profile and the optical function of the key, areal component of the junction. This is was common practice for all devices within the Teddy design that stressed the areal component. Layout and cross-section views can be seen in Figure A.25 whilst key metrics of the devices are shown in table A.28.

Table A.28: Basic features of device TED-3.

TED-3 properties table	
Group membership	H2
Footprint	$500 \times 500\,\mu m$
Basic cell area	$9505.76\,\mu m^2$
Basic cell perimeter	$1020.72\,\mu m$
No. of basic cells	21
Total design area	$199620.96\,\mu m^2$
Total design perimeter	$21435.12\,\mu m$

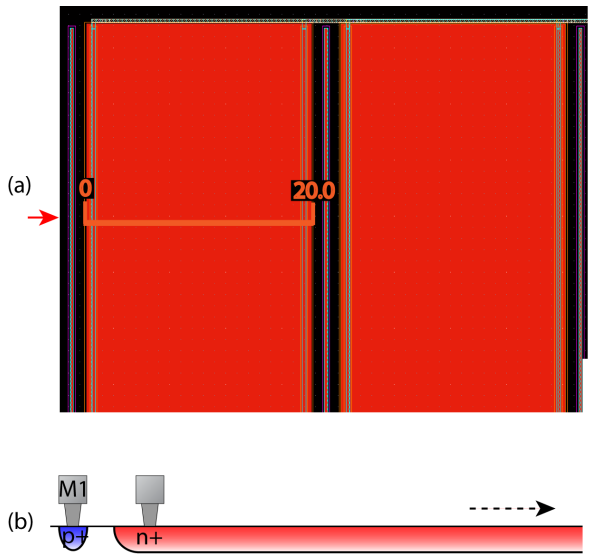

Figure A.25: Part of a pair of basic cells of device TED-3 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a) magnified to show the edge of the basic cell in more detail. (a) The orange bar shows the width of the n+/sub junction. Also visible are three of the substrate anchor contact supply lines that keep the substrate biased at GND in their neighbourhood. (b) Notable features in this cross-section include the substrate anchor and the extent of the well region.

TED-4

TED-4 (Teddy design family, device identifier no.4) is an NW/sub device that is paired to TED-1. Its key characteristic is the comparatively large areal and small side-wall side-wall components. Its 11, gigantic basic cells (in comparison to any other basic cell on any device on any die) span $41 \times 491\,\mu m$ and cover a total of $217977.76\,\mu m^2$ with a corresponding perimeter of $11253.66\,\mu m$. The 41 micron width of each basic cell stretches the capability of the substrate contacts flanking it to maintain an appropriate bias at the entire substrate region underlying the cell, but is still within the safety parameters of the technology. Layout and cross-section views of the device are displayed in Figure A.26 whilst key information appears in table A.29.

Table A.29: Basic features of device TED-4.

TED-4 properties table	
Group membership	H1
Footprint	$500 \times 500\,\mu m$
Basic cell area	$19816.16\,\mu m^2$
Basic cell perimeter	$1023.06\,\mu m$
No. of basic cells	11
Total design area	$217977.76\,\mu m^2$
Total design perimeter	$11253.66\,\mu m$

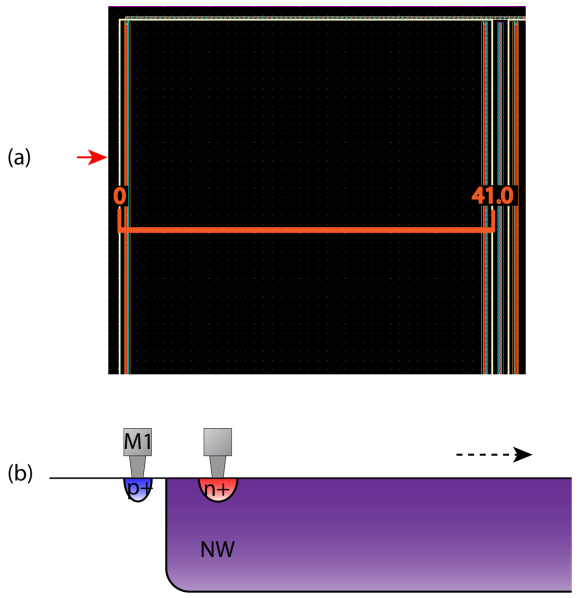

Figure A.26: Part of a pair of basic cells of device TED-4 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a) magnified to show the edge of the basic cell in more detail. (a) The orange bar shows the width of the n+/sub junction. Also visible is one of the substrate anchor contact supply lines that keep the substrate biased at GND in their neighbourhood (far right between the N-well sinkers). (b) Notable features in this cross-section include the substrate anchor and the extent of the well region.

TED-5

TED-5 (Teddy design family, device dientifier no.5) is a nested 3W/NW/sub junction mainly used to study the properties of the UMC13 triple-well. It is paired with TED-9 and acts as the high perimeter and low area counterpart of the pair. It consists of 23472 3W/NW basic cells distributed evenly amongst16 N-well 'tubs'. In total 3W/NW junctions feature an areal component covering $44362.08\,\mu m^2$ with a total perimeter of $129565.44\,\mu m^2$. It's NW/sub component is also worth studying in order to find out how the presence of a 3W region affects the operation of the NW/sub junction as a modulator. The NW/sub component measures $182688\,\mu m^2$ in area and $16640\,\mu m$ in perimeter. TED-5 is also the 5th and final 'large' $500 \times 500\,\mu m$ device. Layout and cross-section views are showcased in Figure A.27 with key information summarised in table A.30.

TED-6

TED-6 (Teddy design family, device identifier no.6) is the only n+/p+ butting diffusion device on Teddy and has no paired device. The geometry of the small device ($300 \times 300\,\mu m$) revolves around thin, long strips of alternating n+ and p+ diffusion regions sitting directly atop substrate. A basic cell is thus defined as a single strip of n+ diffusion forming butting junctions with its p+ neighbours at their side-walls and a regular diffusion-to-substrate junction at the bottom. Thus, 230 basic cells measuring 0.64×294.4 microns were created featuring a total of $42299.3\,\mu m^2$ in area and $132100.2\,\mu m$ in perimeter. Cross-section and layout views can be seen in Figure A.28 whilst important information on the device is summarised in table A.31.

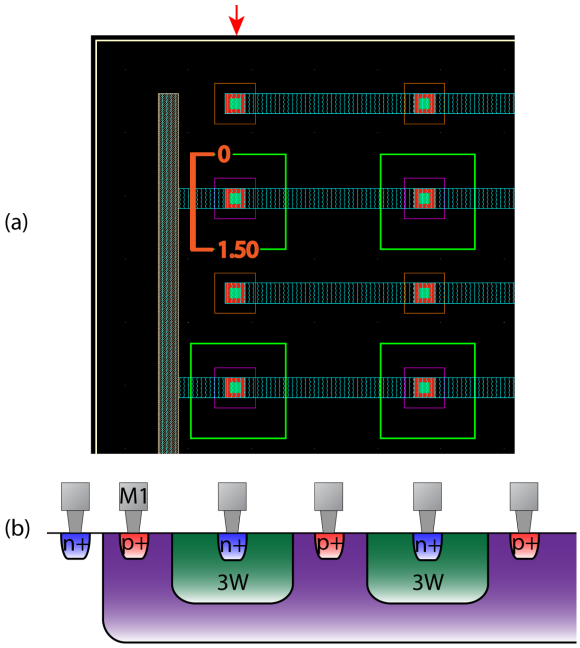

Figure A.27: 3W/NW basic cells of device TED-5 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a) magnified to show the edge of the host NW/sub basic cell in more detail. (a) The orange bar shows the extent of the 3W/NW junction. Also visible are N-well anchor contact supply lines that keep the host N-well biased at a suitable voltage. (b) Notable features in this cross-section include the substrate anchor (beyond the top edge of (a)), the extent of the well region, the N-well sinkers, the extent of the 3W regions and the 3W sinkers linking them to the supply line to their corresponding bond pad.

Table A.30: Basic features of device TED-5.

TED-5 properties table	
Group membership	H3
Footprint	$500 \times 500\,\mu m$
Basic cell area (3W/NW)	$1.89\,\mu m^2$
Basic cell perimeter (3W/NW)	$5.52\,\mu m$
No. of basic cells (3W/NW)	23472
Total design area (3W/NW)	$44362.08\,\mu m^2$
Total design perimeter (3W/NW)	$129565.44\,\mu m$
Basic cell area (NW/sub)	$11.418\,\mu m^2$
Basic cell perimeter (NW/sub)	$1040\,\mu m$
No. of basic cells (NW/sub)	16
Total design area (NW/sub)	$182688\,\mu m^2$
Total design perimeter (NW/sub)	$16640\,\mu m$

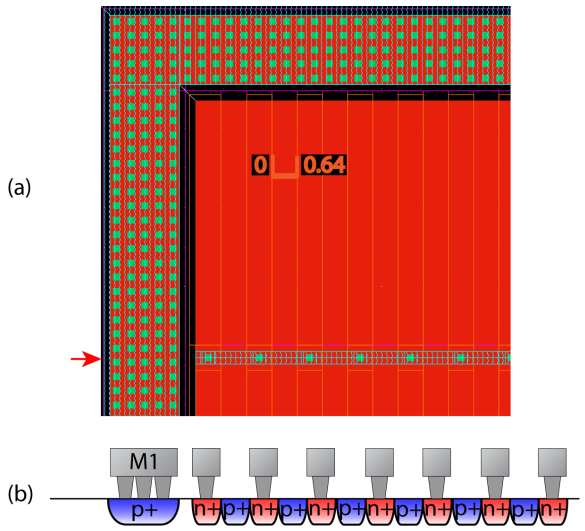

Figure A.28: Part of a basic cell of device TED-6 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bar shows the width of the n+ strips. Also visible is the guard ring of the device. (b) Notable features in this cross-section include the guard ring and the succession of alternating n+ and p+ strips of diffusion.

Table A.31: Basic features of device TED-6.

TED-6 properties table	
Group membership	H5
Footprint	$300 \times 300\,\mu m$
Basic cell area	$183.91\,\mu m^2$
Basic cell perimeter	$575.6\,\mu m$
No. of basic cells	230
Total design area	$42299.3\,\mu m^2$
Total design perimeter	$132100.2\,\mu m$

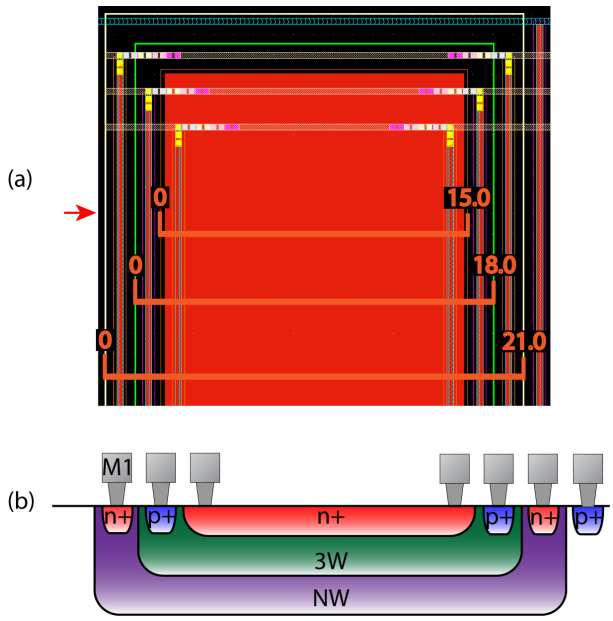

Figure A.29: Part of a basic cell of device TED-7 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show the widths of the n+, 3W and NW strips. Also visible is a substrate anchor supply line at the far right of the image. (b) Notable features in this cross-section include the nested nature of the structure and the positioning of the contacts close to the edges of each individual region.

TED-7

TED-7 (Teddy design family, device identifier no.7) is a nested n+/3W/NW/sub junction device paired with TED-8. Of the pair, TED-7 is the device that consists of basic cells defined by very large areas compared to their perimeters. Thus, 13 basic cells constitute the device and each cell is comprised of an N-well tub measuring $21 \times 290.5\,\mu m$, within which resides a 3W measuring $18 \times 287.5\,\mu m$, within which rests an n+ diffusion region measuring $15 \times 284.5\,\mu m$. The area and perimeter data reflects these sizes and for convenience is simply summarised in table A.32. Cross-sectional and layout views of the device can be found in A.29.

TED-8

TED-8 (Teddy design family, device identifier no.8) is a nested n+/3W/NW/sub device paired with TED-7. Its design revolves around small basic cells of n+/3W residing within a few large N-well tubs, each of which hosts 60 3W islets. Each islet consists of an $8 \times 8\,\mu m$ 3W region that surrounds exactly one $4.46 \times 6\,\mu m$ n+ region. This makes for relatively large total perimeters for the n+/3W and 3W/NW sub-junctions, which stand at $16069\,\mu m$ and $24460.8\,\mu m$ respectively. More detailed information is included in table A.33. Layout and cross-section views can be seen in Figure A.30.

Table A.32: Basic features of device TED-7.

TED-7 properties table	
Group membership	H4
Footprint	$300 \times 300\,\mu m$
Basic cell area (NW/sub)	$5542.70\,\mu m^2$
Basic cell perimeter (NW/sub)	$621.08\,\mu m$
No. of basic cells (NW/sub)	13
Total design area (NW/sub)	$72055.10\,\mu m^2$
Total design perimeter (NW/sub)	$8074.04\,\mu m$
Basic cell area (3W/NW)	$4807.47\,\mu m^2$
Basic cell perimeter (3W/NW)	$609.72\,\mu m$
No. of basic cells (3W/NW)	13
Total design area (3W/NW)	$62497.11\,\mu m^2$
Total design perimeter (3W/NW)	$7926.36\,\mu m$
Basic cell area (n+/3W)	$4085.27\,\mu m^2$
Basic cell perimeter (n+/3W)	$598.36\,\mu m$
No. of basic cells (n+/3W)	13
Total design area (n+/3W)	$53108.51\,\mu m^2$
Total design perimeter (n+/3W)	$7778.68\,\mu m$

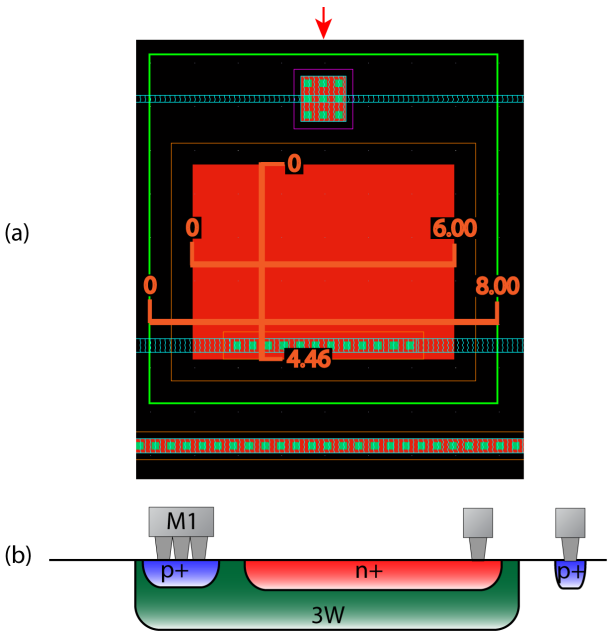

Figure A.30: Part of a basic cell of device TED-8 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show geometrical data on the n+ and 3W regions. Also visible is a substrate anchor supply line at the bottom of the image. (b) Notable features in this cross-section include the nested nature of the structure and the positioning of the contacts.

Table A.33: Basic features of device TED-8.

TED-8 properties table	
Group membership	H4
Footprint	$300 \times 300\,\mu m$
Basic cell area (NW/sub)	$5669.11\,\mu m^2$
Basic cell perimeter (NW/sub)	$605.92\,\mu m$
No. of basic cells (NW/sub)	13
Total design area (NW/sub)	$73698.43\,\mu m^2$
Total design perimeter (NW/sub)	$7876.96\,\mu m$
Basic cell area (3W/NW)	$60.1664\,\mu m^2$
Basic cell perimeter (3W/NW)	$31.36\,\mu m$
No. of basic cells (3W/NW)	780
Total design area (3W/NW)	$46929.79\,\mu m^2$
Total design perimeter (3W/NW)	$24460.80\,\mu m$
Basic cell area (n+/3W)	$25.1152\,\mu m^2$
Basic cell perimeter (n+/3W)	$20.6\,\mu m$
No. of basic cells (n+/3W)	780
Total design area (n+/3W)	$19587.36\,\mu m^2$
Total design perimeter (n+/3W)	$16068.00\,\mu m$

TED-9

The final device of the Teddy family (identifier no.9) is a nested 3W/NW/sub design paired with TED-5. It features larger basic cells than TED-5 and thus represents the 'high area, low perimeter' constituent part of the pair. Notably, because devices TED-5 and TED-9 have as their primary target the study of the 3W structure, the 'host N-wells' in both devices are kept as large as possible in order to host as many 3W 'tubs' as possible. Overall, TED-9 features 780 basic 3W/NW cells that enclose a total area of $24460.80\,\mu m^2$, fenced by $47339.14\,\mu m$ of perimeter. A cross-setional view can be seen along with a layout view in Figure A.31 whilst key metrics of the device are shown in table A.34.

Table A.34: Basic features of device TED-9.

TED-9 properties table	
Group membership	H3
Footprint	$300 \times 300\,\mu m$
Basic cell area (NW/sub)	$5792.99\,\mu m^2$
Basic cell perimeter (NW/sub)	$604.80\,\mu m$
No. of basic cells (NW/sub)	13
Total design area (NW/sub)	$75308.87\,\mu m^2$
Total design perimeter (NW/sub)	$7862.40\,\mu m$
Basic cell area (3W/NW)	$60.6912\,\mu m^2$
Basic cell perimeter (3W/NW)	$31.36\,\mu m$
No. of basic cells (3W/NW)	780
Total design area (3W/NW)	$47339.14\,\mu m^2$
Total design perimeter (3W/NW)	$24460.80\,\mu m$

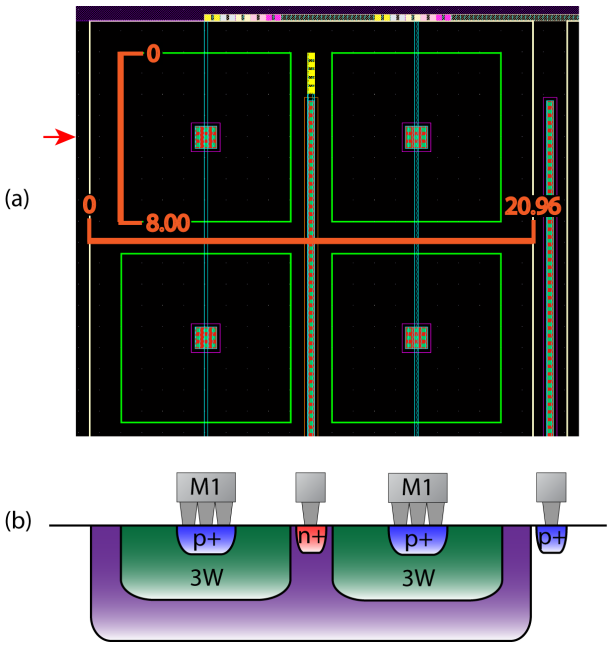

Figure A.31: Basic cells of device TED-9 in (a) top-view and (b) cross-section along the line indicated by the red arrow in (a). (a) The orange bars show geometrical data on the 3W and NW regions. Also visible is a substrate anchor supply line at the far right of the image. (b) Notable features in this cross-section include the nested nature of the structure and the positioning of the 3W, NW and substrate anchor contacts.

A.4 Summary of modulator designs

In summary, the design process generated 29 modulators distributed across three dies, each of which represents a specific technological node. Within each die modulators can be classified by junction type (or types) employed and within sets of devices that employ the same type(s) of junction further subdivided on the basic of features such as geometry, masking etc.

On the basis of the host die, the modulators are divided into those belonging to the Ninja (NIN), the Svejk (SVJ) and the Teddy (TED) design families. The proof-of-concept devices residing on the Ninja are mostly stand-alone structures intended to test for more qualitative differences in modulator performance between various junction types and special features. Thus, for example devices NIN-1 and NIN-5 could be paired into a set of NW/sub junctions that feature different area/perimeter ratios as NIN-5 and NIN-6 could be paired by virtue of sharing both front- and back-end processing exactly with the sole exception of a passivation coating that is present in NIN-5, but absent in NIN-6. On the other hand, the Teddy, being a natural extension to the Ninja hosts devices more geared towards quantitative, comparative results that are meant to be considered within pairs (functional groups) functioning similarly to the NIN-1, NIN-5 and NIN-5, NIN-6 pairs described above. Finally, the Svejk being the last and most advanced of the triplet features devices intended to uncover more subtle aspects of modulator performance, such as non-linear effects that depart from the simple assumption that modulator performance would be a linear combination of the contributions of the areal and side-wall components of each junction. As a result functional groups can be defined for TED and SVJ, but not for NIN.

On the basis of junction type we have worked with all possible pn-junction families that would be

offered by a commercially available triple well-enabled CMOS process with butting diffusion junctions. These are all valid combinations formed by the use of n+, p+, 3W, NW and sub regions. Because of each junction family's specific 'character' we considered devices featuring different junction types as entirely separate entities that were to be examined in isolation from each other.

On the basis of junction geometry we examined devices with different area/perimeter ratios in an attempt to find an empirical formula describing the individual contribution of areal and side-wall components within any pn-junction. Moreover, certain special features were also examined, such as the presence/absence of passivation layer over a modulator and masks that would shade very specific areas of a modulator's surface in order to determine more subtle influences of layout on performance.

Table A.35 summarises the modulator designs developed throughout this project.

Table A.35: Summary of all devices designed within the framework of this project. The A/P ratio expresses the exposed area over exposed perimeter ratio. Special features are an allusion to any other features aside from geometry used to tailor each individual device. In multiple junction devices the sub-junction in parentheses denotes the key sub-junction of the device. A/P data corresponds to that sub-junction. The 'perim mask' label indicates that a mask covering the perimeter of some sub-junction has been used. This is explicitly given except in cases where the answer is obvious.

MODULATOR DESIGN SUMMARY

Device ID	**Junction type**	**Group**	**A/P** (μm)	**Footprint** (μm^2)	**Special feature**
NIN-1	NW/sub	-	1.208	479×479	
NIN-2	(p+/NW)/sub	-	0.183	479×479	
NIN-2	p+/(NW/sub)	-	7.317	479×479	
NIN-4	n+/sub	-	16.742	479×479	
NIN-5	NW/sub	-	3.660	479×479	
NIN-6	NW/sub	-	3.660	479×479	No passivation
NIN-7	(n+/p+)/NW	-	0.600	479×479	Butting
NIN-8	n+/p+	-	0.251	479×479	
NIN-9	n+/sub	-	0.259	479×479	Poly
SVJ-1	n+/sub	G1	∞	300×300	Perim mask
SVJ-2	n+/sub	G1	∞	300×300	Perim mask
SVJ-3	(p+/NW)/sub	G2	1.816	300×300	NW perim mask
SVJ-3	p+/(NW/sub)	-	∞	300×300	NW perim mask
SVJ-4	(p+/NW)/sub	G2	∞	300×300	NW & p+ perim mask
SVJ-4	p+/(NW/sub)	-	∞	300×300	NW & p+ perim mask
SVJ-5	n+/sub	G3	1.562	300×300	
SVJ-6	n+/sub	G3	∞	300×300	Perim mask
SVJ-7	NW/sub	G4	0.316	200×200	
SVJ-8	NW/sub	G4	0.455	200×200	Rectangular cell
SVJ-9	NW/sub	G4	0.973	200×200	
SVJ-10	NW/sub	G4	0.973	200×200	Rectangular cell
SVJ-11	NW/sub	G4	2.567	200×200	
SVJ-12	(3W/NW)/sub	G5	0.596	200×200	
SVJ-12	3W/(NW/sub)	-	5.882	200×200	
TED-1	NW/sub	H1	0.264	495×495	
TED-2	n+/sub	H2	0.285	495×495	
TED-3	n+/sub	H2	9.313	495×495	
TED-4	NW/sub	H1	19.369	495×495	
TED-5	(3W/NW)/sub	H3	0.342	495×495	
TED-5	3W/(NW/sub)	-	10.979	495×495	
TED-6	n+/p+/sub	H5	0.320	299×299	Butting
TED-7	(n+/3W)/NW/sub	H4	6.827	299×299	
TED-7	n+/(3W/NW)/sub	-	7.885	299×299	
TED-7	n+/3W/(NW/sub)	-	8.924	299×299	
TED-8	(n+/3W)/NW/sub	H4	1.219	299×299	
TED-8	n+/(3W/NW)/sub	-	1.919	299×299	
TED-8	n+/3W/(NW/sub)	-	9.356	299×299	
TED-9	(3W/NW)/sub	H3	1.935	299×299	
TED-9	3W/(NW/sub)	-	9.579	299×299	

Appendix B

Electro-optical modulation: raw data and supporting experiments

This part of the appendix contains raw results obtained from measuring our devices under the test bench and test protocols described in zappendix 5.4. Results from the tests ran with the described test-bench were first obtained in a 'raw' format (straight out of the measurement device) and then more 'processed' versions (extracting differences, ratios etc.) were developed.

The current appendix module is organised thusly: Section B.1 will tackle the obtained results both in terms of the numerical values obtained, but also in terms of their meaning and interpretation within the context of the test set-up. Certain supporting experiments aimed at placing the obtained results into perspective will be summarised in section B.2.

B.1 Measured results

In this section an explanation of the result processing methodology is presented, followed by raw results of modulator performance testing on all devices. An important note concerning nomenclature is that in this chapter data is classified into the following categories:

- Raw data: This is experimentally obtained data, such as measurements of photocurrent during the alignment phase of the modulation test procedure.
- Design data: Any parameter extracted from the design of the modulator will fall into this category; for example the size of the areal junction of a device or its junction type.
- Derived data: Information that arises as a result of processing raw data and possibly combining it with design data, both originating from a single device.
- Relational data: Any result that arises from the combination of data from different devices.

In terms of result processing, experimentally obtained information undergoes a two-tier process throughout its transformation from raw data to conclusions. The first processing step concerns the

aggregation of all obtained raw data into tables and then the extraction of some derived results and will be tackled in this section. First tier data thus maximises the understanding of a single device given the experimentally obtained results and readily available design data. The second step concerns the combination of raw, design and derived data between experiments or experimental sessions and offers a deeper insight into the operation of our devices in the quality of modulators. More specifically, second tier data concerns more elaborate derived data and relational data oriented towards uncovering relations between devices and dies (includes trend extraction). These issues are considered in subsequent sections.

Note: Where multi-junction devices are concerned test are ran in such way that any piece of semiconductor not directly related to the junction under study is left floating or grounded. The exception is the substrate that is invariably held at GND. Whether a region of semiconductor is to be grounded or left floating depends on the following rule: the side of the test-junction physically closest to the substrate is always the one that is grounded. Anything between that region and the substrate (i.e. 'downstream') is also grounded. Any regions 'upstream' are left floating. Thus, for example when testing the 3N junction of a n3NS device, the n-region is left floating, biasing is applied on the 3-region, the N-region is grounded and S is grounded by default. When testing the n3 junction of the same device, the n-region receives the bias voltage and all other regions remain grounded. The decision to leave the upstream regions floating was related to practicality as the set-up did not allow for shorting upstream connections without major modifications. This issue is discussed in more detail in subsection 5.6.3.1.

Note: All results in this section have been taken while using the same light bulb. This may help consistency, yet prolonged use of the light bulb does lead to performance degradation. This issue is considered farther down this chapter.

B.1.1 Result processing methodology

In order to assess modulation, paired tests with and without illumination were conducted according to the protocols presented in section 5.4.6. This yielded pairs of A_{p-p} scalars; one value for pure pick-up (no illumination) and one for pick up plus optical effects (full illumination). Even though the obtained A_{p-p} values constituted the core data, the set-up of the oscilloscope was arranged in such way that the minimum and maximum singleton peak-to-peak values recorded during the constitution of A_{p-p} could also be noted along with the standard deviation within the collection of singleton values constituting A_{p-p}. These would allow some insight into the statistical significance of the results (by use of the z-test for example). Along with the alignment current, the aforementioned collected results represent all the raw data recorded from any individual device. Transient signal or FFT data recorded in parallel proved of little use because of the inherent weakness of the signal and has thus only been taken within the context of supporting experiments. Table B.1 shows a typical completed table of raw values for a single device.

The morsels of raw data described so far have been taken for every device of each die tested, however certain devices on SVJ2 (the 2nd test device of the Svejk design) have been tested in both set-up I and II, and while in set-up I, certain extra pieces of information were collected. These are values of system output when only a certain 'standard amount' of pick-up is present both under conditions of full illumination and in the dark. By standard amount it is meant that within a rigid test set-up the pick-up was engineered to be as close to constant as possible regardless of which device is being measured.

Table B.1: Example of a full table of raw results from a single device on a single die (in this case NIN1-1). L_{on}: Illumination present. L_{off}: illumination absent. Top row: description of value (Illum.: Illumination indicator (on or off). Min: minimum singleton peak-to-peak value, Max: maximum singleton peak-to-peak value, Align: alignment current.) Second row: units.

	Illum.	A_{p-p}	**Min**	**Max**	**Deviation**	**Align**
		mV	mV	mV	mV	μA
NIN1-1	L_{on}	354.2	349	359	2.3	165
	L_{off}	352	347	355	1.7	

Physically this was achieved by turning all the switches connecting devices of the test SVJ die to the instrumentation to standard 'neutral' positions. This immediately implies the following: a) All devices are disconnected from the signal generator so the pick-up that invades the system can only be due to charge accumulation/depletion on PCB tracks and connectors. b) Because the position of the switches is standard, the PCB tracks and connectors that cause the pick-up noise are always the same when the PCB is in this 'neutral' state.

The reason why neutral measurement pairs (illuminated and dark) have to be taken for each individual device (a seemingly counter-intuitive endeavour) is that for each device, the neutral measurement pair reveals the effect of illumination on the photodetector when the said device is in full alignment; i.e. under the specific illumination conditions that prevail while the test device is in full alignment. More information about the significance of this approach can be found in section 5.5.1. As an example a table illustrating the complete set of data for device SVJ2-7 in set-up I is given in B.2. This can be compared to the largely similar table in B.1. Note: the standard deviation of the neutral values for all SVJ-type devices tested for this metric were always in the region of $0.2 - 0.5\,mV$ with some exceptions[1].

Table B.2: Example of a full table of raw results from a single device on a single die (in this case SVJ2-7) in set-up I featuring the extra information on pick-up in the neutral state (see text). Top row: description of value (Min: minimum singleton peak-to-peak value, Max: maximum singleton peak-to-peak value, Align: alignment current. Neutral: value corresponding to 'Avg', but taken in the neutral state of the set-up.) Second row: units. L_{on}: Illumination present. L_{off}: illumination absent.

	A_{p-p}	**Min**	**Max**	**Deviation**	**Align**	**Neutral**
	mV	mV	mV	mV	μA	mV
SVJ2-7 L_{on}	15.39	14.8	15.9	0.25	0.868	11.73
SVJ2-7 L_{off}	15.61	15.1	16.3	0.19		12.1

Tier 1 data that can then be extracted from the raw data will include: the absolute difference between illuminated and dark A_{p-p} values, their fractional difference, their 2-sample z-scores and whether there is any overlap in the range of A_S values recorded. This derived data, along with the original raw data and key design data can be combined to provide a comprehensive summary of the device. For those devices where data on testing in the neutral state is available one may also calculate the fractional difference between the $8\,kHz$ components under illumination and in the dark respectively in the neutral state and compare it to the corresponding A_{p-p} fractional difference. Then, based on the neutral state fractional

[1]During some of the tests an unknown source of noise increased the standard deviation to approx $1.5\,mV$. The noise source was never determined, but it was external to the set-up as it took effect at seemingly random times and on different devices each time.

difference and A_{p-p} in the dark a predicted value for A_{p-p} under illumination can be extracted and compared to the measured value. The difference between predicted and measured values may potentially provide a measure of optical modulation effects if certain assumptions are found to hold true. These possibilities will be further discussed in a different section (5.6). An example taken from SVJ2-7 when tested in set-up I with the supplementary neutral state information illustrates what a collection of tier-1 data looks like (table B.3).

Note: fractional error 'F' throughout this chapter will be understood as the value yielded by the formula $F = \frac{a-b}{a+b}$ i.e. difference over average. The measure is chosen as a more advanced form of $F = a/b$ despite the weak point when it comes to handling value pairs with average equal to zero. This is done in order to achieve a magnitude-commutative relation as exchanging a and b will merely change the sign of F and not its magnitude.

Table B.3: Tier-1 results for device SVJ2-7. DEV: Device identifier. Type: Junction type. B_{p-p}: Absolute difference between results under illumination and in darkness (light ON - light OFF). N/F %: Fractional difference between results under illumination and in the dark (light ON / light OFF). N/F (n): Same as N/F, but in the neutral state. Z-score: The statistical z-score of the result pair (illuminated and dark). Align: Alignment photocurrent of the device. Ovrlp: Overlap between ranges of A_S values (Yes or No). Area: Areal junction extent. Side: Side-wall junction extent. The middle row holds units, where applicable. Z-scores are measured in units of pooled standard errors, which we simply call σ_E.

DEV	**Type**	B_{p-p}	**N/F**	**N/F (n)**	**Z-score**	**Align**	**Ovrlp**	**Area**	**Side**
		mV	%	%	σ_E	μA		μm^2	μm
SVJ2-7	NS	-0.22	98.59	96.94	4.95	0.868	Y	7045	22282

Throughout the rest of this section raw and tier-1 results will be presented from each individual die, sorted by design (i.e. all dies of the Svejk design grouped together etc.). A short section of comments is also provided for each die, whereby a few observations are made on the available data. Thorough investigation of the data-set, however, is reserved for section 5.5.

B.1.2 The 'Ninja' family performance

Only one specimen of the Ninja family has been processed and all tests performed on it have been carried out in configuration II.

NIN1

Raw and tier-1 results are shown in tables B.4 and B.5 respectively. Devices numbers 7 and 8 failed to show rectifying characteristics and were therefore considered failed. No further results were obtained from them after failing their diagnostic tests.

Comments: The raw results show that regardless of device, there is always a very strong background pick-up component atop which differences due to illumination ride. Both A_{p-p} values for fully illuminated and dark condition experiment runs tend to be of the order of $10^3\,mV$ for all devices involved whilst the difference between them is of the order of $10^0\,mV$ or at most $10^1\,mV$. Values of standard deviation match well with recorded maxima and minima in the various collected A_S values, and tend to compare

Table B.4: Raw results from die NIN1. Defective devices numbers 7 and 8 have been omitted.

DEV	**Illum.**	A_{p-p}	**Min**	**Max**	**Deviation**	**Align**
		mV	mV	mV	mV	μA
NIN1-1	L_{on}	354.2	349	359	2.3	165
	L_{off}	352.0	347	355	1.7	
NIN1-2-NS	L_{on}	558.3	554	563	1.8	124
	L_{off}	556.0	552	559	1.6	
NIN1-2-pN	L_{on}	165.9	164	169	1.1	2.86
	L_{off}	159.8	159	161	0.7	
NIN1-4	L_{on}	453.9	450	459	1.8	67.5
	L_{off}	427.8	425	431	1.4	
NIN1-5	L_{on}	336.1	333	339	1.6	158
	L_{off}	346.3	343	351	1.3	
NIN1-6	L_{on}	256.2	254	259	1.4	149
	L_{off}	260.6	258	264	1.3	
NIN1-9	L_{on}	460.7	457	467	2.0	35.0
	L_{off}	451.8	449	455	1.5	

Table B.5: Tier-1 results from die NIN1. Defective devices numbers 7 and 8 have been omitted.

DEV	**Type**	B_{p-p}	**N/F**	**Z-scores**	**Align**	**Over**	**Area**	**Side**
		mV	%	σ_E	μA		μm^2	μm
NIN1-1	NS	2.2	100.63	5.44	165	Y	75204	62280
NIN1-2-NS	NS	2.3	100.41	6.75	124	Y	84142	11500
NIN1-2-pN	pN	6.1	103.82	33.08	2.86	N	27878	152064
NIN1-4	nS	26.1	106.10	80.93	67.5	N	206626	12342
NIN1-5	NS	-10.2	97.05	34.99	158	N	133848	36566
NIN1-6 (PAD)	NS	-4.4	98.31	16.29	149	Y	133848	36566
NIN1-9	nS	8.9	101.97	25.17	35.0	N	78766	304134

in size with the differences between illuminated and dark A_{p-p} cases. Finally, alignment currents seem to be making a strong differentiation between well and diffusion-based devices. Well-based devices are generally far more efficient at capturing optical power and transforming it into electrical form. This is a subject of study in a different chapter (chapter 6). Here, photocurrents are only used for alignment purposes.

Tier-1 results are largely a quantified version of the observations above. Apart from that, the yielded z-scores indicate that there is a very high probability that the different values obtained for illuminated and dark experiment runs feature a statistically significant difference, although they offer no clue on the origins of the said difference.

B.1.3 The 'Svejk' family performance

In total, two specimens of Svejk were tested, namely Svejk 1 and Svejk 2. Of them, Svejk 2 was selected for testing under set-up I as well as the standard set-up II.

SVJ1 - configuration II

Results can be seen in tables B.6 (raw) and B.7 (tier-1).

Comments: Raw results indicate that strong pick-up seems to set the baseline with light-based changes occurring on top. These changes are statistically significant since the gap between results under illumination and in the dark is much wider than the standard deviation and even the maximum and minimum values recorded in each case cover non-overlapping intervals.

Tier-1 results confirm the observations above by the large (triple-digit) values of z-scores obtained for every device. Moreover, with the notable exception of SVJ1-12-NS the result for an illuminated set-up yield significantly lower A_{p-p} values that in the dark case.

SVJ2 - configuration II

Results are presented in tables B.8 (raw) and B.9 (tier-1).

Comments: Both raw and tier-1 results point in the same direction as those for SVJ1, including the SVJ-12-NS device behaving in seemingly the opposite way to all other devices (A_{p-p} under illumination higher than in the dark).

SVJ2 - configuration I

Results are presented in tables B.10 (raw) and B.11 (tier-1).

Comments: Raw results immediately show that baseline activity in configuration I is far lower than in configuration II. Single- and double-digit baseline activities are the norm as opposed to triple-digit seen in configuration II. Most standard deviations of A_{p-p} are sub-mV, again in contrast to configuration II. Interestingly, levels of 'pure pick-up' (neutral state) seem fairly consistent for all pn-junctions with

Table B.6: Raw results from die SVJ1 in test configuration II

DEV	**Illum.**	A_{p-p}	**Min**	**Max**	**Deviation**	**Align**
		mV	mV	mV	mV	μA
SVJ1-1	L_{on}	1252.0	1245	1258	2.56	0.68
	L_{off}	1351.1	1346	1356	2.56	
SVJ1-2	L_{on}	1779.2	1770	1780	3.54	0.77
	L_{off}	1884.3	1880	1890	4.92	
SVJ1-3-NS	L_{on}	270.7	268	274	1.31	13.05
	L_{off}	410.3	407	413	1.57	
SVJ1-3-pN	L_{on}	169.5	166	173	1.37	0.90
	L_{off}	450.8	448	454	1.56	
SVJ1-3-pNS	L_{on}	157.2	154	160	1.47	13.8
	L_{off}	251.6	249	255	1.36	
SVJ1-4-NS	L_{on}	271.1	268	275	1.9	9.50
	L_{off}	413.3	411	417	1.36	
SVJ1-4-pN	L_{on}	163.0	160	166	1.46	0.22
	L_{off}	451.2	448	454	1.58	
SVJ1-4-pNS	L_{on}	177.6	173	182	1.75	9.30
	L_{off}	270.4	267	273	1.13	
SVJ1-5	L_{on}	170.2	167	175	1.74	2.86
	L_{off}	263.5	261	268	1.30	
SVJ1-6	L_{on}	182.1	180	186	1.38	1.54
	L_{off}	292.7	289	296	1.52	
SVJ1-7	L_{on}	947.4	943	950	1.69	8.12
	L_{off}	1024.3	1020	1027	1.56	
SVJ1-8	L_{on}	279.0	274	282	1.55	12.03
	L_{off}	382.9	380	386	1.40	
SVJ1-9	L_{on}	149.1	146	152	1.35	13.73
	L_{off}	245.1	243	248	1.16	
SVJ1-10	L_{on}	201.6	198	205	1.62	12.81
	L_{off}	283.3	280	287	1.33	
SVJ1-11	L_{on}	168.2	166	172	1.32	15.77
	L_{off}	254.3	252	257	1.19	
SVJ1-12-NS	L_{on}	908.5	905	913	1.67	13.55
	L_{off}	683.1	680	686	1.57	
SVJ1-12-3N	L_{on}	693.6	690	698	1.89	0.52
	L_{off}	752.2	747	755	1.72	
SVJ1-12-3NS	L_{on}	260.2	257	265	1.67	13.5
	L_{off}	320.6	317	323	1.36	

Table B.7: Tier-1 results from die SVJ1 in test configuration II. Where junctions have been measured in parallel unified area and side-wall measurements have no meaning.

DEV	Type	B_{p-p}	N/F	Z-scores	Align	Over	Area	Side
		mV	%	σ_E	μA		μm^2	μm
SVJ1-1	nS	-99.1	92.67	193.55	0.68	N	11615	0
SVJ1-2	nS	-105.1	94.42	122.61	0.77	N	11615	0
SVJ1-3-NS	NS	-139.6	65.98	482.76	13.05	N	31819	0
SVJ1-3-pN	pN	-281.3	37.60	958.06	0.9	N	30628	16865
SVJ1-3-pNS	pNS	-94.4	62.48	333.32	13.8	N	N/A	N/A
SVJ1-4-NS	NS	-142.2	65.59	430.33	9.5	N	6757	0
SVJ1-4-pN	pN	-288.2	36.13	947.29	0.22	N	6757	0
SVJ1-4-pNS	pNS	-92.8	65.68	315.01	9.3	N	N/A	N/A
SVJ1-5	nS	-93.3	64.59	303.74	2.86	N	38698	24768
SVJ1-6	nS	-110.6	62.21	380.94	1.54	N	21976	0
SVJ1-7	NS	-76.9	92.49	236.43	8.12	N	7045	22282
SVJ1-8	NS	-103.9	72.86	351.75	12.03	N	10615	23316
SVJ1-9	NS	-96	60.83	381.38	13.73	N	16017	16466
SVJ1-10	NS	-81.7	71.16	275.62	12.81	N	10615	23316
SVJ1-11	NS	-86.1	66.14	342.57	15.77	N	24524	9554
SVJ1-12-NS	NS	225.4	133.00	695.35	13.55	N	24559	4175
SVJ1-12-3N	3N	-58.6	92.21	162.15	0.52	N	13844	23232
SVJ1-12-3NS	3NS	-60.4	81.16	198.30	13.5	N	N/A	N/A

the notable exception of those that are nested inside a well (pN, 3N). There is no obvious reason for such behaviour.

Tier-1 results show fairly statistically significant results have been obtained for the larger part (z-scores generally over 2σ) although the overlap of the ranges of individual A_S values seems to indicate that these differences are very small. Interestingly, the comparison between 'N/F' and 'N/F (n)' columns shows that there are some discrepancies between the ratios of A_{p-p} under illumination vs. in the dark for the modulator devices and the reference pick-up signals. It is not obvious from the data though whether this is significant or reveals a trend.

B.1.4 The 'Teddy' family performance

In total two samples of the Teddy design family were tested; samples number 5 and 6. Samples 1 through 4 failed at various stages of the packaging process and were discarded. All specimens of the Teddy family were tested in configuration II alone.

TED5

Results for the 5th die of the Teddy design are displayed in tables B.12 and B.13 for raw and tier-1 results respectively.

Comments: On first examination of the raw results TED5, much like NIN1, exhibits a large baseline pick-up component and small variations riding atop it. These variations tend to be comparable, but generally larger than the standard deviation of each population of A_S values. TED5-8-NS and TED5-

Table B.8: Raw results from die SVJ2 in test configuration II.

DEV	**Illum.**	A_{p-p}	**Min**	**Max**	**Deviation**	**Align**
		mV	mV	mV	mV	μA
SVJ2-1	L_{on}	1185.8	1181	1191	1.96	0.72
	L_{off}	1273.4	1270	1277	1.69	
SVJ2-2	L_{on}	1764.1	1760	1770	4.89	0.82
	L_{off}	1855.1	1850	1860	5.04	
SVJ2-3-NS	L_{on}	258.4	254	262	1.66	12.05
	L_{off}	320.7	318	324	1.2	
SVJ2-3-pN	L_{on}	196.8	194	200	1.48	0.92
	L_{off}	404.1	402	407	1.3	
SVJ2-3-pNS	L_{on}	201.5	197	206	2.24	12.1
	L_{off}	232	230	236	1.41	
SVJ2-4-NS	L_{on}	254.6	251	257	1.5	7.5
	L_{off}	316	313	320	1.55	
SVJ2-4-pN	L_{on}	183.8	181	188	1.5	0.21
	L_{off}	399.4	396	403	1.42	
SVJ2-4-pNS	L_{on}	207	204	211	1.55	7.1
	L_{off}	238.9	236	241	1.25	
SVJ2-5	L_{on}	161.8	159	166	1.61	3.0
	L_{off}	221.7	219	225	1.18	
SVJ2-6	L_{on}	166.7	162	174	2.25	1.53
	L_{off}	231.3	229	234	1.2	
SVJ2-7	L_{on}	816.5	811	821	2.34	8.13
	L_{off}	856.1	853	860	1.82	
SVJ2-8	L_{on}	119.9	117	123	1.32	12.28
	L_{off}	183.5	181	186	1.12	
SVJ2-9	L_{on}	138.9	136	143	1.37	13.36
	L_{off}	185.2	182	188	1.28	
SVJ12-10	L_{on}	171.1	167	175	1.61	12.11
	L_{off}	213.8	211	218	1.32	
SVJ-2-11	L_{on}	151.4	148	154	1.42	15.0
	L_{off}	194	190	197	1.18	
SVJ2-12-NS	L_{on}	634.9	630	640	2.24	13.1
	L_{off}	456	452	460	1.53	
SVJ2-12-3N	L_{on}	581.7	577	586	1.78	0.51
	L_{off}	623	620	626	1.44	
SVJ2-12-3NS	L_{on}	149.9	146	154	1.81	14.1
	L_{off}	179.1	176	182	1.34	

Table B.9: Tier-1 results from die SVJ2 in test configuration II.

DEV	Type	B_{p-p}	N/F	Z-scores	Align	Over	Area	Side
		mV	%	σ_E	μA		μm^2	μm
SVJ2-1	nS	-87.6	93.12	239.35	0.72	N	11615	0
SVJ2-2	nS	-91	95.09	91.63	0.82	N	11615	0
SVJ2-3-NS	NS	-62.3	80.57	215.07	12.05	N	31819	0
SVJ2-3-pN	pN	-207.3	48.7	744.13	0.92	N	30628	16865
SVJ2-3-pNP	pNS	-30.5	86.85	81.48	12.1	N	N/A	N/A
SVJ2-4-NS	NS	-61.4	80.57	201.28	7.5	N	6757	0
SVJ2-4-pN	pN	-215.6	46.02	738.08	0.21	N	6757	0
SVJ2-4-pNS	pNS	-31.9	86.65	113.28	7.1	N	N/A	N/A
SVJ2-5	nS	-59.9	72.98	212.19	3	N	38698	24768
SVJ2-6	nS	-64.6	72.07	179.13	1.53	N	21976	0
SVJ2-7	NS	-39.6	95.37	94.46	8.13	N	7045	22282
SVJ2-8	NS	-63.6	65.34	259.78	12.28	N	10615	23316
SVJ2-9	NS	-46.3	75.00	174.62	13.36	N	16017	16466
SVJ2-10	NS	-42.7	80.03	145.03	12.11	N	10615	23316
SVJ2-11	NS	-42.6	78.04	163.15	15	N	24524	9554
SVJ2-12-NS	NS	178.9	139.23	466.34	13.1	N	24559	4175
SVJ2-12-3N	3N	-41.3	93.37	127.55	0.51	N	13844	23232
SVJ2-12-3NS	3NS	-29.2	83.70	91.68	14.2	N	N/A	N/A

5-NS along with TED5-5-3N to a lesser extent differentiate themselves from the rest of the data by exhibiting extraordinarily high baseline figures (TED5-5-NS greatly exceeding every other device on the die).

Tier-1 results offer a number of observations, including the fact that 6/14 junctions show a greater A_{p-p} in the dark, while the remaining 8/14 exhibit the opposite. The ON/OFF A_{p-p} ratios also tend the be rather small ($< 10\%$ ON-OFF difference) yet despite this, all differences in A_{p-p} bar TED5-7-NS seem to be unequivocally statistically significant according to the z-score values. Note: TED5-7-NS is a device that exhibits particularly weak change in A_{p-p} with illumination.

TED6

Results for the 5th die of the Teddy design are displayed in tables B.14 and B.15 for raw and tier-1 results respectively.

Comments: Raw results show similar outcomes to TED5. Devices TED6-8-NS, TED6-5-NS and TED6-5-3N show extraordinarily high baseline figures with TED6-5-NS exhibiting by far the highest.

Tier-1 results, similar to TED5, show a mixture of negative (5/14) and positive (9/14) B_{p-p} values and relatively small differences between illuminated and dark A_{p-p} results. In fact, 5/14 junctions exhibit z-scores in the single digits, the minimum being 2.39 for TED6-5-3N.

Table B.10: Raw results from SVJ2 in test configuration I.

DEV	**Illum.**	A_{p-p}	**Min**	**Max**	**Deviation**	**Align**	**Neutral**
		mV	mV	mV	mV	μA	mV
SVJ2-1	L_{on}	16.81	15.9	17.9	0.45	0.047	12.45
	L_{off}	17.42	16.4	18.3	0.42		12.65
SVJ2-2	L_{on}	17.86	16.7	18.4	0.36	0.051	11.91
	L_{off}	18.54	17.9	19.0	0.29		12.35
SVJ2-3-NS	L_{on}	14.35	13.4	15.4	0.50	0.870	12.37
	L_{off}	14.97	13.6	15.9	0.46		12.67
SVJ2-3-pN	L_{on}	11.21	9.0	15.9	1.22	0.143	10.28
	L_{off}	10.50	7.6	14.8	1.65		10.12
SVJ2-3-pNS	L_{on}	12.56	12.0	13.1	0.27	1.430	11.93
	L_{off}	12.84	12.4	13.3	0.20		12.14
SVJ2-4-NS	L_{on}	14.99	14.1	15.5	0.30	0.547	11.96
	L_{off}	15.48	14.9	16.0	0.23		12.29
SVJ2-4-pN	L_{on}	7.63	6.9	9.1	0.45	0.031	5.86
	L_{off}	8.04	7.0	8.9	0.46		6.13
SVJ2-4-pNS	L_{on}	13.03	12.5	14.2	0.24	0.855	12.32
	L_{off}	13.17	12.7	14.4	0.34		12.55
SVJ2-5	L_{on}	13.08	12.8	13.7	0.21	0.306	11.93
	L_{off}	13.30	13.0	13.8	0.25		12.16
SVJ2-6	L_{on}	12.43	11.6	13.4	0.38	0.161	11.12
	L_{off}	12.84	12.2	13.8	0.25		11.36
SVJ2-7	L_{on}	15.19	14.6	15.7	0.29	0.879	11.39
	L_{off}	15.56	15.0	16.1	0.20		11.82
SVJ2-8	L_{on}	12.73	12.1	13.2	0.22	1.305	11.51
	L_{off}	13.31	12.9	14.3	0.32		11.87
SVJ2-9	L_{on}	13.07	12.5	13.6	0.20	1.426	10.97
	L_{off}	13.57	13.0	14.7	0.37		11.04
SVJ2-10	L_{on}	13.25	12.9	14.0	0.26	1.320	11.70
	L_{off}	13.71	13.0	14.6	0.36		11.98
SVJ2-11	L_{on}	13.03	12.8	13.8	0.19	1.553	11.90
	L_{off}	13.18	12.9	14.3	0.25		12.20
SVJ2-12-NS	L_{on}	15.70	15.4	16.4	0.21	1.470	11.91
	L_{off}	15.75	15.4	16.4	0.23		12.11
SVJ2-12-3N	L_{on}	9.44	8.6	10.4	0.39	0.063	6.16
	L_{off}	9.81	8.8	11.1	0.41		5.55
SVJ2-12-3NS	L_{on}	13.52	13.0	15.1	0.38	1.555	12.15
	L_{off}	13.80	13.0	14.7	0.34		12.23

Table B.11: Tier-1 results from SVJ2 in configuration I. The 'N/F (n)' column shows the ratio of A_{p-p} for the 'reference pick-up' or 'neutral' configuration (illuminated/dark).

DEV	**Type**	B_{p-p}	**N/F**	**N/F (n)**	**Z-scores**	**Align**	**Over**	**Area**	**Side**
		mV	%	%	σ_E	μA		μm^2	μm
SVJ1-1	nS	-0.61	96.50	98.42	7.01	0.047	Y	11615	0
SVJ1-2	nS	-0.68	96.33	96.44	10.40	0.051	Y	11615	0
SVJ1-3-NS	nS	-0.62	95.86	97.63	6.45	0.870	Y	31819	0
SVJ1-3-pN	pN	0.71	106.76	101.58	2.45	0.143	Y	30628	16865
SVJ1-3-pNS	pNS	-0.28	97.82	98.27	5.89	1.430	Y	N/A	N/A
SVJ1-4-NS	nS	-0.49	96.83	97.31	9.17	0.547	Y	6757	0
SVJ1-4-pN	pN	-0.41	94.90	95.6	4.51	0.031	Y	6757	0
SVJ1-4-pNS	pNS	-0.14	98.94	98.17	2.38	0.855	Y	N/A	N/A
SVJ1-5	nS	-0.22	98.35	98.11	4.76	0.306	Y	38698	24768
SVJ1-6	nS	-0.41	96.81	97.89	6.37	0.161	Y	21976	0
SVJ1-7	nS	-0.37	97.62	96.36	7.43	0.879	Y	7045	22282
SVJ1-8	nS	-0.58	95.64	96.97	10.56	1.305	Y	10615	23316
SVJ1-9	nS	-0.5	96.32	99.37	8.41	1.426	Y	16017	16466
SVJ1-10	nS	-0.46	96.64	97.66	7.32	1.320	Y	10615	23316
SVJ1-11	nS	-0.15	98.86	97.54	3.38	1.553	Y	24524	9554
SVJ1-12-NS	nS	-0.05	99.68	98.35	1.14	1.470	Y	24559	4175
SVJ1-12-3N	3N	-0.37	96.23	110.99	4.62	0.063	Y	13844	23232
SVJ1-12-3NS	3NS	-0.28	97.97	99.35	3.88	1.555	Y	N/A	N/A

B.2 Supporting experiments

On occasion the necessity appeared, to create a new test protocol and apply it to a new test-bench in order to provide supplementary information that would help understand the results of the main modulator testing. These additional tests are individually presented in this section, from test-bench set-up to results. The aggregation of information yielded by these experiments with data from the modulator performance experiments is carried out in the next section.

B.2.1 Beam power measurement

This experiment was intended to determine how much optical power is emitted from the collimator outlet. The beam exiting the collimator is neither entirely consistent of parallel beams nor uniform in intensity, yet a very large component of the beam falls entirely within the bounds of the test IC die (set-up I) or within the bounds of the combined IC die and photodetector active areas (set-up II). It is important to know how much power reaches the photodetector in order to understand how the reference beam power affects modulator performance.

It could be argued that what really matters is how much optical power crosses through the modulator, which does indeed highly impact the modulation capabilities of any device. However, due to the imperfections in the test set-ups (both I and II), all we know is that a certain amount of light (impossible to determine quantitatively under the set-ups used), part of which will be modulated (again impossible to determine exactly how much) will hit the photodetector. Thus, by taking a worst-case scenario we can say that one needs at least X amount of power in order to run a modulator and obtain the results presented in this chapter. If in some future set-up the electro-optical components can be arranged in such way that all the light reaching the photodetector is modulated, then system performance can only

Table B.12: Raw results from TED5 in test configuration II.

DEV	Illum.	A_{p-p}	Min	Max	Deviation	Align
		mV	mV	mV	mV	μA
TED5-1	L_{on}	280.1	275	286	2.5	93.6
	L_{off}	267.7	264	271	1.8	
TED5-2	L_{on}	244.5	239	250	2.3	10.6
	L_{off}	239.4	236	244	1.8	
TED5-3	L_{on}	271.7	265	279	2.5	8.80
	L_{off}	264.5	261	268	1.5	
TED5-4	L_{on}	270.1	265	276	2.6	91.3
	L_{off}	295.6	290	301	1.7	
TED5-7-n3	L_{on}	115.4	114	119	1.1	0.84
	L_{off}	118.5	117	120	0.7	
TED5-7-3N	L_{on}	361.5	357	367	2.0	1.80
	L_{off}	347.1	343	351	1.7	
TED5-7-NS	L_{on}	473.0	467	480	2.2	33.5
	L_{off}	472.6	470	476	1.5	
TED5-8-n3	L_{on}	159.8	157	162	1.2	0.39
	L_{off}	167.4	166	169	0.6	
TED5-8-3N	L_{on}	379.4	375	388	2.1	1.91
	L_{off}	359.3	356	363	1.5	
TED5-8-NS	L_{on}	791.9	787	798	2.6	43.4
	L_{off}	801.6	798	807	1.7	
TED5-5-NS	L_{on}	2688	2680	2690	3.6	83.0
	L_{off}	2726	2660	2730	9.3	
TED5-5-3N	L_{on}	629.1	613	633	2.8	3.52
	L_{off}	621.4	618	627	1.7	
TED5-9-NS	L_{on}	491.3	488	495	1.9	37.4
	L_{off}	507.8	502	511	1.4	
TED5-9-3N	L_{on}	301.5	298	305	1.5	2.25
	L_{off}	282.9	281	286	1.0	

Table B.13: Tier-1 results from TED5 in test configuration II.

DEV	Type	B_{p-p}	N/F	Z-scores	Alignment	Over	Area	Side
		mV	%	σ_E	μA		μm^2	μm
TED5-1	NS	12.4	104.63	28.46	93.6	N	35123	133004
TED5-2	nS	5.1	102.13	12.35	10.6	Y	127116	446332
TED5-3	nS	7.2	102.72	17.46	8.8	Y	199621	21435
TED5-4	NS	-25.5	91.37	58.04	91.3	N	217978	11254
TED5-7-n3	n3	-3.1	97.38	16.81	0.84	Y	53109	7779
TED5-7-3N	3N	14.4	104.15	38.79	1.8	N	62497	7926
TED5-7-NS	NS	0.4	100.08	1.06	33.5	Y	72055	8074
TED5-8-n3	n3	-7.6	95.46	40.06	0.39	N	19587	16068
TED5-8-3N	3N	20.1	105.59	55.07	1.91	N	46930	24461
TED5-8-NS	NS	-9.7	98.79	22.08	43.4	Y	73698	7877
TED5-5-NS	NS	-38	98.61	26.94	83.0	Y	182688	16640
TED5-5-3N	3N	7.7	101.24	16.62	3.52	Y	44362	129565
TED5-9-NS	NS	-16.5	96.75	49.44	37.4	N	75308	7862
TED5-9-3N	3N	18.6	106.57	72.96	2.25	N	47339	24461

Table B.14: Raw results from TED6 in test configuration II.

DEV	Illum.	A_{p-p}	Min	Max	Deviation	Align
		mV	mV	mV	mV	μA
TED6-1	L_{on}	374.5	370	381	2.4	90.7
	L_{off}	344.5	337	350	2.2	
TED6-2	L_{on}	364.3	360	370	2.1	10.3
	L_{off}	327.7	324	332	1.7	
TED6-3	L_{on}	385.4	381	394	2.5	8.5
	L_{off}	347.0	343	351	2	
TED6-4	L_{on}	357.9	352	365	2.8	87.8
	L_{off}	359.6	355	363	1.8	
TED6-7-n3	L_{on}	212.3	197	219	-	0.82
	L_{off}	191.6	188	195	1.8	
TED6-7-3N	L_{on}	348	343	355	3.2	1.72
	L_{off}	344.7	339	352	2.8	
TED6-7-NS	L_{on}	529.7	526	534	1.6	31.9
	L_{off}	519.1	516	522	1.8	
TED6-8-n3	L_{on}	168.1	166	171	1.1	0.389
	L_{off}	175.3	174	176	0.5	
TED6-8-3N	L_{on}	353.7	348	360	2.2	1.87
	L_{off}	349.8	346	353	1.9	
TED6-8-NS	L_{on}	770.9	765	777	2.2	41.4
	L_{off}	795.0	792	800	1.8	
TED6-5-NS	L_{on}	2675.6	2670	2680	5.0	80.2
	L_{off}	2719.8	2710	2720	1.3	
TED6-5-3N	L_{on}	609.6	605	617	2.3	3.4
	L_{off}	608.7	603	611	1.5	
TED6-9-NS	L_{on}	457.8	465	471	1.6	36.4
	L_{off}	493.4	490	496	1.4	
TED6-9-3N	L_{on}	270.5	268	274	1.4	2.23
	L_{off}	262.8	261	264	0.6	

Table B.15: Tier-1 results from TED6 in test configuration II.

DEV	Type	B_{p-p}	N/F	Z-scores	Alignment	Over	Area	Side
			%	σ_E	μA			
TED6-1	NS	30	108.71	65.16	90.7	N	35123	133004
TED6-2	nS	36.6	111.17	95.79	10.3	N	127116	446332
TED6-3	nS	38.4	111.07	84.81	8.50	N	199621	21435
TED6-4	NS	-1.7	99.53	3.61	87.8	Y	217978	11254
TED6-5-n3	n3	20.7	110.80	81.32	0.84	N	53109	7779
TED6-5-3N	3N	3.3	100.96	5.49	1.72	Y	62497	7926
TED6-5-NP	NS	10.6	102.04	31.12	31.9	N	72055	8074
TED6-8-n3	n3	-7.2	95.89	42.13	0.39	N	19587	16068
TED6-8-3N	3N	3.9	101.11	9.49	1.87	Y	46930	24461
TED6-8-NP	NS	-24.1	96.97	59.95	41.4	N	73698	7877
TED6-5-NP	NS	-44.2	98.37	60.50	80.2	N	182688	16640
TED6-5-3N	3N	0.9	100.15	2.32	3.40	Y	44362	129565
TED6-9-NP	NS	-35.6	92.78	118.40	36.4	N	75308	7862
TED6-9-3N	3N	7.7	102.93	35.75	2.23	N	47339	24461

improve. As a result, the experiment described in this subsection carries certain value as a 'lower bound' estimate of the power needs of our proposed electro-optical modulators.

Additionally, tests can be ran whereby the actual amount of light reaching the photodetector after crossing through dies manufactured in different technologies and hosting different devices can be measured in order to provide some perspective in how adding a modulator die affects the net illumination at the photodetector. This is useful when it comes to calculating what we can characterise as 'steady state' beam losses, i.e. beam losses that occur beyond the control of the engineer.

Set-up and procedure: This experiment consists of two parts. The first part consists of determining an upper bound of the maximum power that can be extracted from the photodetector under full illumination with no modulator chip in between. This yields an indirect measure of the reference beam power. The second part comprises of tests whereby the short-circuit current of the photodetector is measured with different dies blocking some of the light arriving from the collimator and then the obtained values are compared to a maximum photocurrent test run without any modulator die involved.

For part 1: The photodetector was positioned at an elevation (distance form the collimator outlet) where the collimated beam was at its highest optical density. The photodetector was then aligned for maximum photocurrent. The maximum photocurrent value was taken, then the SMU was switched to open circuit mode and the open-circuit voltage was recorded. Multiplying the two yields a simplistic upper bound on the electrical power that can be harvested from the reference beam.

For part 2: The photodetector was positioned at a convenient elevation and aligned for maximum photocurrent with no modulator die covering it (set-up II). Subsequently, and with no change in alignment dies were added over the photodetector one by one and the resulting photocurrents measured. The ratios of photocurrents obtained through modulator dies over the reference photocurrent obtained without a modulator die can offer some insight into losses suffered through the different die families.

Results: In part 1 the results were a short-circuit current of magnitude $820\,\mu A$ and an open circuit voltage of $195.5\,mV$ yielding total electrical power of $160.31\,\mu W$ of electrical power. This value cannot be directly linked to the optical power leaving the collimator, but it can be said that whatever the illumination scheme, if the spectral content is similar and a photodetector of the same type as the one we used can provide approx. $160\,\mu W$ of electrical power, then the results of our experiments should be repeatable. Of course, with more efficient illumination schemes (spectral content-wise) or photodetectors, this indicative, lower-bound value may well drop significantly. Further study is required in order to confirm or disprove this hypothesis.

The values obtained in part 2 are shown in table B.16. Note: in the packaging implementation chosen (section 5.4.3) the dies do not cover the entire extent of the hole that provides optical access to the photodetector and therefore some light may reach the photodetector while entirely by-passing the modulator dies. Moreover differences in the degree of coverage by metallisation between dies will further distort results. For that reason the values shown in table B.16 are imperfect and need to be considered with caution. They do, however show that a fair amount of light lying in the active wavelength range of the photodetector ($800 - 1800\,nm$) is stopped by the Silicon dies.

Table B.16: Effect of inserting different dies in the optical path of the reference beam. Values are all in micro-amps and represent photodetector zero-bias photocurrents. The column marked 'Ref' shows the photocurrent obtained when no die lies between the photodetector and the collimator.

	Ref	**NIN**	**TED**	**SVJ**
I_{photo}	227	40.7	90.9	83.6

B.2.2 Beam profile specification

The profile of the beam exiting the collimator was found not to be uniform. For that reason an experiment was devised whereby the profile of the beam was roughly determined.

Set-up and procedure: A $25 \times 25 \mu m$ photodiode was placed on the motorised positioner at an elevation above the positioner platform level that corresponds roughly to the position of the modulator die in set-up II. The step size of the positioner is approximately 100 microns. Photocurrent was measured from the photodiode. The origin (0,0) point was arbitrarily set to the rough positioner location where the photocurrent was globally maximum, then excursions up to 700 microns in the 'x-direction' and 1mm in the 'y-direction' were imposed upon the photodiode and photocurrent measurements were taken after every 100 micron step. The x- and y-directions were chosen arbitrarily at the start of all experiments and remained consistent throughout. These measurements, being qualitative in nature, concerned themselves with fractional changes in photocurrent rather than absolute values.

Results: Plots of normalised photocurrent vs. location can be seen in Figure B.1. Notably, both data sets seem to have a broad, single peak which implies that during the modulator alignment process that single peak is the the only valid reference point to which all devices are aligned. Moreover, given that no device is larger than $500 \times 500\,\mu m$ it would seem that the beam profile is largely constant (within 5% of peak amplitude at most). Finally, neither x-direction nor y-direction profiles are perfectly smooth. Particularly in the y-direction there seems to be a small region of locally reduced optical power output at -200 microns from the origin. This implies that simply interpolating, or even fitting, between sample points may not be enough to determine the exact nature of the beam profile thus introducing more uncertainty in any attempts to evaluate beam power within a certain radius of the peak irradiance point.

Given the above three observations and the knowledge that no set-up is perfectly rigid, thus leading to position uncertainty, we shall simply assume that regardless of device size, the beam profile is uniform and qualitatively the same. This will introduce certain errors when attempting to compare data such as A_{p-p} per unit area between devices of different footprint sizes. Positioning uncertainty is impossible to assess quantitatively with a good degree of accuracy but it can be said that it depends on simple mechanical components of the set-up such as the screws holding the PCB stacks together and attach to holes in the positioner platform, all of which experience a certain degree of movement.

B.2.3 Light source noise evaluation

When running experiments according to the protocols described in section 5.4.6 some early test-runs were specifically ran in order to determine the effects of light source noise on the set-up within the test context. Also, short-term repeatability tests were ran within the context of this experiment.

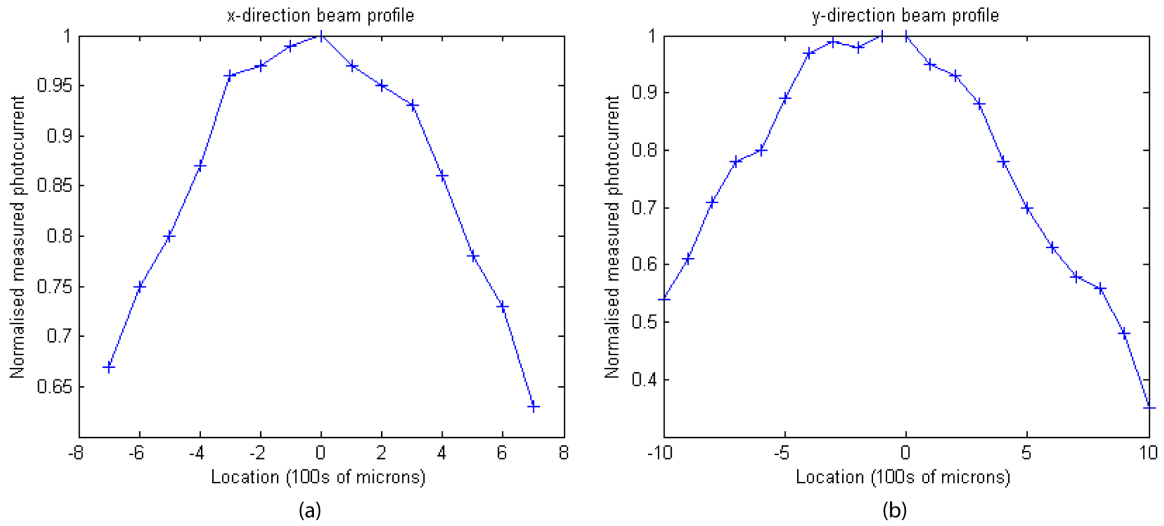

Figure B.1: x- (a) and y-direction (b) measured beam profiles for electro-optical modulation experiments. The origins of each plot represent the same reference point. Plots are normalised to the signal amplitude at the origin (selected to represent the point of peak amplitude).

Set-up and procedure: The experimental set-up was unchanged with respect to the main procedures described in section 5.4.6, but the nature of the measurements was. Overall a set of five measurements was taken during each round in a strictly defined order:

- Light on, signal on: Same protocol as the 'full illumination' protocol in the main experimental measurement runs.
- Light off, signal on: Same protocol as the 'dark' protocol in the main experimental measurement runs.
- Light off, signal off: This additional protocol was intended to reveal the A_{p-p} value in the absence of both light and signal, thus yielding a 'baseline' value for comparison with all other measurement.
- Light on, signal off: Yields an A_{p-p} value with just the illumination on. Intended to reveal the effects of the light source alone on the result.
- Repeat of light on, signal on: Repeat of the first measurement of the round. Intended to reveal short term drift.

Each round of measurements (all 5 runs) lasted approximately five minutes overall.

Despite the fact that the entire TED5 and TED6 dies were tested in this fashion, only some results from TED5 will be shown as indicative results (TED5-1, TED5-2, TED5-3, TED5-4, TED-6-n3). Similar results were obtained from all other devices residing on TED5 and TED6 and are not included here. The shown devices were selected simply for being the very first devices tested this way in the TED die series. A_{p-p} values are given only even though maximum, minimum and standard deviation values are also available.

Results: Results from this experiment can be split into raw and tier-1 results. These can be seen in tables B.17 and B.18 respectively.

Table B.17: Raw results for the light-source noise determination experiment. Legend: L_{on}: light on. L_{off}: light off. S_{on}: signal on. S_{off}: signal off. All units are in mV.

State	A_{p-p}				
	TED5-1	TED5-2	TED5-3	TED5-4	TED5-7-n3
L_{on}, S_{on}	245.8	208.2	249.8	244.0	146.7
L_{off}, S_{on}	249.8	222.1	262.7	284.1	145.5
L_{off}, S_{off}	14.0	14.9	16.1	13.3	12.4
L_{on}, S_{off}	15.6	16.4	19.8	15.2	13.9
L_{on}, S_{on}	243.4	225.1	252.7	244.0	146.8

Table B.18: Tier-1 results for the light source noise determination experiment. Legend is similar to table B.17. Ratios are taken as light on over light off ratios. Differences accordingly as light on minus light off.

State	**Tier-1 results**				
	TED5-1	TED5-2	TED5-3	TED5-4	TED5-7-n3
S_{on} ratio	0.98	0.94	0.95	0.86	1.01
S_{off} ratio	1.11	1.1	1.23	1.14	1.12
S_{on} difference	-4	-13.9	-12.9	-40.1	1.2
S_{off} difference	1.6	1.5	3.7	1.9	1.5

Raw data shows that indeed most of the strength of A_{p-p} depends on the presence of pick-up from the signal generator. Also a quick scan of the table reveals that the differences introduced by the light-source state are far larger when the signal generator is active than when it is inactive.

Tier-1 results quantify the above observations. Indeed it would seem that the light source tends to reduce the amplitude of the A_{p-p} when the signal generator is active and increase it slightly when it is inactive. This could be a result of electro-optical modulation, but it could also be because the light source may affect the sensitivity of the photodetector with regards to pick-up (reduce it) while at the same time adding noise.

Should the latter be the case, then at low noise amplitudes (no pick-up) the noise addition prevails whilst under conditions of strong pick-up the reduction in sensitivity masks and buries the added noise. The underlying mechanism could be a change in small-signal impedance at the photodetector, brought about by the sudden influx of photocurrent. More details will be given in section 5.6 as this effect could dramatically affect core results. Note: The exact contribution of light source noise to A_{p-p} varies between devices (see table B.18) and between homologous devices[2] on TED5 and TED6 (results not shown). Thus quantifying it is not a trivial effort.

Finally, in terms of short-term drift, we see that comparing the first and last measurement run within a round, A_{p-p} values that are meant to be identical are generally not. In some cases the discrepancies are rather large (extreme case: TED5-2), whilst in other cases they are small (extreme case: TED5-4). The

[2]Homologous devices: devices that form different physical 'instances' of a common design blueprint. For example, the set Q = {SVJ1-1, SVJ2-1, SVJ3-1 } forms a set of homologous devices matching the pattern SVJ-1 and residing on SVJ1, SVJ2 and SVJ3.

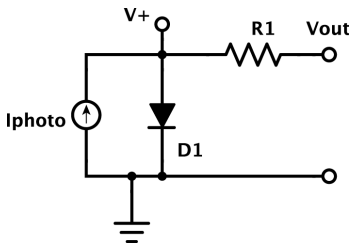

Figure B.2: Simplified macromodel of a photodetector. R1: shunting resistance, I_{photo}: photocurrent.

reasons for these discrepancies are unknown and can vary between interference from other instruments in the anechoic chamber where the experiments were conducted and fast thermal effects.

Additional considerations: A number of issues arise that may distort the results to a certain extent with one in particular standing out. Examining the succession of measurements it is seen that the light source spends one run on, two off, and then two on again. When multiple rounds are ran back-to-back the light stays on between rounds. Thus the effective succession is three runs plus the round set-up time on and two runs off. The light source is based on a halogen light bulb and therefore will be subject to heating when on and cooling immediately after switch-off. This may affect reference beam output and therefore corrupt results. Generally, observations of photocurrent as a function of time with the light source 'cold-started' (switched on for the first time in the day of observation) and the rest of the set-up still show that changes in photocurrent are small (a few percentage points) and slow (15' time frame). As an example the photocurrent through TED5-1 dropped from $24,1\,\mu A$ to $23.65\,\mu A$ after 15', $23.48\,\mu A$ after 30' and $23.38\,\mu A$ after 2hrs and 30', signifying a drop of approx. 3% over the entire 2hrs 30'. Thus this effect has been discounted from our conclusions.

Results shown in table B.17 do not match those shown in table B.12. The reason for this may be a combination of long-term drift and the fact that due to catastrophic failure of the light-bulb in the time interval between the measurements that led to table B.17 and those that led to B.12, the two sets of results were taken with different light bulbs inside the illuminator.

B.2.4 Photodetector light-dependent anode-cathode voltage variation testing

The macromodel of a photo-electric element consists of a diode, a photocurrent source, a shunt resistor, a parasitic series resistor and the junction capacitance. Here we shall concentrate on DC behaviour under different levels of background illumination which means a simplified macromodel consisting of the diode, photocurrent source and parasitic series resistor will suffice. The shunting resistor is considered too large to significantly affect results (Figure B.2).

Suppose the terminals of the photodetector are shorted. Under no illumination both voltage nodes in the system remain at GND potential and the small signal resistance of the diode is the 0-bias resistance R_0. Assuming we apply a small signal V_{pert} at node V_+, the 'disturbance current' it will create will equal $V_{pert}/(R_0||R_1)$.

Under a certain amount of illumination a photocurrent starts to flow through the system. As the photocurrent drains to GND via the parallel combination of R_0 and series resistor R_1 a voltage difference

starts to develop across $V1-GND$ and the values of R_0 begins to change. The system reaches equilibrium at the point where the photocurrent is entirely drained to GND via the parallel combination of R_{Vx} and R_1, where R_{Vx} is the small-signal resistance of the diode. Applying a small signal disturbance to node V_+ will now result in a disturbance current of $V_{pert}/(R_{Vx}||R_1)$.

In conclusion the same small-signal perturbation at node V_+ will cause different disturbance currents to flow towards the shorted terminals of the photodetector. This is important in the case where the photodetector is connected to an ideal current-sensing amplifier. The terminals of the device will both be held at virtual earth and the disturbance current is what the amplifier will sense. This result is also generally valid, albeit with added complications, even if the amplifier is not ideal.

The intention of this experiment is to determine whether under the photodetector-side experimental set-up this phenomenon does occur to a measurable level or not.

Set-up and procedure: The connection of instruments follows the standard protocol used for modulation, configuration II. The BNC cable that connects the photodetector to the preamplifier is split into two cables, a long ($1.2\,m$) and a short one ($0.3\,m$), connected by a T-junction. The orientation of this assembly plays a significant role in the experiment. Thus we distinguish between the photodetector to long BNC to T-junction to short BNC to preamplifier (long-short) and the photodetector to short BNC to T-junction to long BNC to preamplifier (short-long) configurations.

The core measurements consist of SMU readings taken out of the T-junction 'tap' along the transmission line consisting of the two BNC cables either side of it. The measurements are taken either under full illumination or in the dark and concern the DC voltage value observed at the tap. Thus frequency-dependent transmission line effects can be safely ignored with only the resistive component of the line playing any significant role. Results are taken for both long-short and short-long configurations.

At the same time the standard A_{p-p} values are monitored as a means of checking whether modulator performance is affected. In order to provide better comparative results extra A_{p-p} measurements are taken: a) performance with a regular connector, as used in standard modulation tests. b) Performance with the T-junction 'extra' terminal unconnected. The device used for testing was SVJ2-7.

Results: Results are shown in table B.19. Of particular note are the following facts: a) The connectivity seems to hardly at all affect A_{p-p} results. The presence of light, however does under all circumstances. b) The connectivity does affect the value read out by the SMU at the T-junction 'tap'. The difference may be small, but still indicates that the effect is occurring. Unfortunately the 'true voltage' at the terminals of the photodetector is not as readily accessible and cannot be determined. However, since between the tap and the photodetector lies a BNC cable, a PCB connector and the PCB tracks it is safe to assume that the voltage difference with respect to GND will be even higher at the terminals of the photoetector; how much greater, is unknown.

Notes: The $-4.9\,mV$ value recorded commonly for both long-short and short-long configurations seems to be the 'virtual earth' of the preamplifier. Moreover, in order to test for light source drift (possibly due to thermal effects) the values read by the SMU were left to stabilise. However, apart from a drop of the short-long value under illumination from $1.4\,mV$ to $1.3\,mV$ after about 5' no other changes were observed with time.

Additional: In order to understand how this change in bias voltage will affect small-signal impedance

Table B.19: Photodetector light-dependent anode-cathode voltage variation testing results.

Connectivity	**Conditions**	A_{p-p}	**Spot voltage**
		mV	mV
Regular connector	L_{on}	795.8	
	L_{off}	845.8	
T-junction floating	L_{on}	794.0	
	L_{off}	844.5	
Short-long	L_{on}	795.8	1.4
	L_{off}	846.9	-4.9
Long-short	L_{on}	794.1	0.4
	L_{off}	845.7	-4.9

we revisit the governing relation between current and voltage in an ideal diode:

$$I = I_s \left(e^{V/nV_T} - 1 \right) \tag{B.1}$$

where I is the diode current, I_s the reverse saturation current; a quantity that is related to the geometry of the device, V is the cross-terminal voltage and V_T the thermal voltage; a quantity that is approximately $25\,mV$ at room temperature. n is the diode ideality factor, which for simplicity is assumed to be equal to 1 in our case.

Equivalently, the voltage as a function of current is given by:

$$V = V_T \ln\left(\frac{I + I_s}{I_s}\right) = V_T \ln(I + I_s) - V_T \ln(I_s) \tag{B.2}$$

Consequently the impedance can be expressed as:

$$Z(V|I) = \frac{dV}{dI} = \frac{V_T}{I + I_s} = \frac{V_T}{I_s e^{V/V_T}} \tag{B.3}$$

where $Z(V|I)$ is the impedance of the diode as a function of either voltage or current.

Thus we can compare impedances at a given reference level V_0 and a 'reference plus perturbation' level $V_0 + V_p$:

$$Q = \frac{Z(V_0)}{Z(V_0 + V_p)} = \frac{e^{\frac{V_0 + V_p}{V_T}}}{e^{\frac{V_0}{V_T}}} = e^{\frac{V_p}{V_T}} \tag{B.4}$$

where Q is the ratio of impedances at the baseline and baseline plus perturbation level.

What equation B.4 effectively states is that so long as equation B.1 holds, any perturbation V_p in the voltage across the terminals of the device will lead to a fractional impedance change with a coefficient of $\exp(V_p/V_T)$; that is to say the exponential of the perturbation voltage in units of V_T. For example even a relatively small increase of $10\,mV$ at the voltage across the terminals of a diode will lead to a roughly 50% increase in impedance if the thermal voltage is assumed to be equal to $25\,mV$.

B.2.5 System noise specification

A basic diagnostic test was ran in order to determine noise levels in the overall system in both configurations I and II. These tests were ran primarily in order to determine noise bottlenecks.

Set-up and procedure: Standard set-up I and II configurations were employed, but this time the oscilloscope was set to display an FFT (Fast Fourier Transform) of the raw transient signal coming from the preamplifier rather than a locked-in version. The noise floor was measured in the dark and under full illumination with no instruments other than the essentials (light source, preamplifier, oscilloscope) switched on. Then all instruments were switched on under full illumination and the noise measurement was repeated.

Noise-floor values were taken at above $2\,kHz$ frequencies where the noise spectral density was approximately flat. Spikes were detected a certain frequencies. Those frequencies were always consistent and were avoided by use of the standard, $8\,kHz$ measurement protocol when taking the main results.

Results: Results can be seen in table B.20. Results seem to indicate that the instruments themselves seem to introduce little white noise in the system; below the clearly discernible level. Moreover, in configuration I, where the light source is farther from the photodetector, a significantly reduced portion of the light reaches the photodetector.

Table B.20: System noise values under different test conditions and configurations.

Noise level (dBV)	**Test configuration**	
Conditions	**I**	**II**
L_{off}	-115	-115
L_{on}	-115	-110
Instruments on	-115	-110

B.2.6 Effects of filtering light

A long-pass filter with $1300\,nm$ cut-off wavelength was available. As the light source used was broad-band with the main components into the visible range it was deemed interesting to see whether cutting off frequencies below $1300\,nm$ might offer an improved SNR (Signal to Noise Ratio). This would theoretically occur because the intensity of the free-carrier absorption phenomenon depends on the square of the wavelength but the light source-induced noise may simply depend on the total irradiance provided to the photodetector adjusted for the spectral sensitivity of the photodetector (this subject requires further study). According to this hypothesis there is a possibility that the SNR at each narrow wavelength band increases as the band wavelengths involved become longer. Thus, cutting off all visible and 'low-grade' NIR light could possibly leave the more useful high wavelength components continue towards their modulator target alone and relatively unabated.

Set-up and procedure: The test used standard configuration II for modulation testing, but the oscilloscope traced the raw, rather than the locked-in version of the input signal. DC coupling was employed and the average voltage level over $2\,s$ was recored over $50 \times 2\,s$ cycles. The 'average of averages' (the G_{p-p}) over the 50 cycles was the reference value meant to represent the 'true' DC level while the asso-

ciated standard deviation would provide an indirect estimate of noise magnitude. Overall three 50-cycle test runs were carried out: one for unfiltered, full illumination, one for 1300 nm long-passed illumination and one in the dark as a reference of general background noise. This test protocol was employed since the exact value of the standard deviation is irrelevant; only differences between values obtained through application of a consistent methodology were important. Moreover, consistency of the experimental protocol used for this test with the main result extraction framework was deemed as potentially useful. Note: The time resolution of the oscilloscope was considered as sufficient at 200 ms/div for the purposes of our experiment.

A secondary objective of the test was to see whether the G_{p-p} would differ when different bias voltages were applied to the modulator device. The DUT (device under test) was NIN1-1. A test was run at 0-bias and another at $2\,V$ reverse bias.

Results: The results can be seen in table B.21. It would seem that indeed the noise added by radiation under 1300 nm is disproportionately large compared to the increase in baseline activity. Nevertheless, the standard deviation values themselves come with a lot of uncertainty and cannot simply be taken at face value. Moreover, the full illumination, $2\,V$ reverse bias standard deviation ($660\,\mu V$) is unusually high. Such outliers are not uncommon although generally extracted standard deviation values from the application of the $50 \times 2\,s$ averages protocol tend to be reasonably repeatable, to within maybe 30% (from observations made over the course of time).

Moreover, no significant differences were found between results obtained at different modulator device bias levels. Given variations in the light source output even if the results obtained were proven to be statistically significant to a large degree the premise that what is being observed is actual amplitude modulation of incoming light would still be questionable.

Note: these are rough, back of the envelope results. We could use the data given in table B.21 to attempt to calculate specifically how much noise is added to the system by radiation over the 1300 nm cut-off and radiation below it, but given the qualitative nature of the result we are after and the complexities arising from dealing with the uncertainty factors associated with this experiment we propose that the precise intricacies of the SNR vs light wavelength band should be the subject of future studies.

Table B.21: Measurements of DC signal level and variation under different conditions of illumination and modulator reverse bias. STD stands for standard deviation.

Status	**0V modulator bias**		**2V modulator bias**	
	DC level	**STD**	**DC level**	**STD**
	mV	μV	mV	μV
No FEL	365.58	390	365.42	660
With FEL	172.57	140	172.57	220
Dark	6.62	90	6.63	80

B.2.7 Performance creep

Performance creep is what we will call long-term modulator performance variation that cannot be attributed to any obvious change in experimental protocol or experiment conditions. We can only speculate as to what causes performance creep and repeatedly measure A_{p-p} values of certain, reference devices

with a difference of a few days in order to attempt and quantify the phenomenon. For this experiment every device on TED5 was tested with a difference of at least 4 days between trials. Note: these trials were performed with a different lamp compared to the main results of this chapter (subsection B.1) so differences between these auxiliary and the main results can be at least partially attributed to the change in lamp.

Set-up and procedures: The set-up of this experiment is the standard configuration II used for extracting the main results and the experimental protocols used are the same as the main experimental protocol. This was in order to ensure that no change in either test configuration or protocol could be the source of trial-to-trial variation. Provided results are all with full illumination. Similar changes were observed in the dark (results not shown).

Results: At the end of the experiment we obtained the table of results shown in B.22. Only A_{p-p} results from six representative devices are shown for brevity. Results for all devices were largely similar in quality.

Table B.22: A_{p-p} value pairs obtained from six different devices on TED5 with a time difference of at least 4 days. The ratios of the 1st run values over the 2nd run values are also given.

DEV	**TED5-1**	**TED5-2**	**TED5-4**	**TED5-5-NS**	**TED5-3N**	**TED5-8-n3**
1st run	245.8	208.2	244.0	2692.0	624.1	180.7
2nd run	248.1	237.4	254.0	2697.0	610.0	207.7
Ratio	0.991	0.877	0.961	0.998	1.023	0.870

The obtained results show that trial-to-trial variation can be quite large. As for the causes of this phenomenon one can speculate that they would include differences in temperature between the days when the measurements were taken, failure to align the devices to the light outlet in exactly the same way each time, light bulb performance decay and ambient electrical noise of unspecified origins.

Appendix C

PMU subcircuit simulated results

Throughout its design phase the power management unit had to be subjected to various simulations both in terms of individual parts and in more complex modules and as a whole system. In this chapter of he appendix we examine a number of important tests in a logical order that gradually builds up from the 'starting point' of the system, the ring oscillator, all the way to the full system.

In order to test for the fitness of purpose of our design we chose a simplistic nominal model whereby an ideal voltage source equal to $0.5\,V$ acts as the unregulated power supply from the power harvester units. Measuring the current drawn from the voltage source at any given time and taking the maximum value we can obtain an estimate of how much current needs to be diverted from the optical scavengers to the PMU and circuits farther downstream. In subsequent tests the intention was to replace the $0.5\,V$ voltage source with a realistic photodetector macromodel as described in chapter 4. In retrospect, a much better simulation strategy would be to use a photodiode macromodel from the beginning. Its constant current source could be set to a realistically achievable photocurrent value given the geometrical specifications of our devices and based, even if loosely, on experimentally measured data from optical harvesters in similar technologies (devices manufactured in IBM18 were not readily available until the fabricated SVEJK die arrived).

Note: quoted power dissipation measurements are typically extracted by integrating the current drawn from a constant power supply over one microsecond and then computing the average power dissipation within that microsecond. Thus the values given are not free of error. For example, in tests ran on the clock generator the choice of window limits may affect the result of the current integration, particularly when the system is running at low frequency.

Digital frequency data is drawn manually by examining a 'typical' cycle in the transient plot and computing the difference between two periodically homologous points (e.g. square waveform 'corners') in order to obtain the waveform period. In order to compute duty cycles the timestamps of two, specially selected, consecutive data-points in the waveform are subtracted and compared to the period. The said points must: a) contain the 'high' stage of the waveform and b) have a voltage value at least approximately equal to VDD/2. These manual estimates are not perfect and thus small errors exist in frequency and duty cycle ratio values as well.

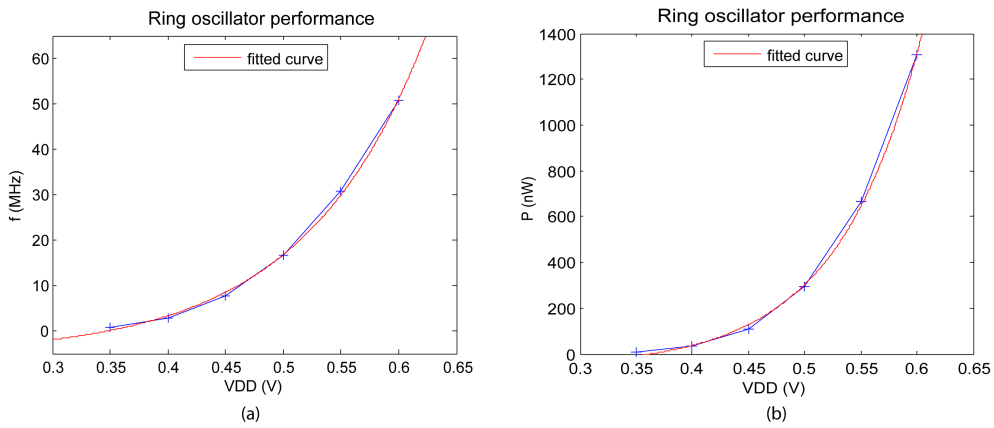

Figure C.1: Frequency vs. VDD (a) and power dissipation vs. VDD (b) curves for the ring oscillator. Crosses represent data points whilst red, smooth lines represent the exponential fits.

C.1 Ring oscillator

The ring oscillator is a very simple system that receives a power supply and generates a regular digital waveform whose frequency depends on the voltage supply and the current allowed to enter it. The ring oscillator implemented in this PMU did not employ any current-starved techniques, which means that it possesses a non-tunable frequency vs. voltage supply characteristic curve. In terms of power dissipation we may plot a power vs. voltage supply curve that will link a range of possible sustained solar cell voltage outputs to the power drawn by the ring oscillator. Plots of frequency and power dissipation vs. voltage supply can be seen in Figure C.1. The corresponding table of results is C.1. Note: At nominal $0.5\,V$ supply the obtained values for frequency and power dissipation are $16.68\,MHz$ and $293.7\,nW$ respectively.

Table C.1: Key metrics extracted from the ring oscillator: output frequency and power dissipation as a function of supply voltage.

RING OSCILLATOR METRICS		
VDD	f_{clk}	**P**
V	MHz	nW
0.35	0.82	8.18
0.40	2.90	32.46
0.45	7.63	108.50
0.50	16.68	293.70
0.55	30.77	663.30
0.60	50.76	1306.20

The frequency vs. power supply plot has been fitted to an exponential with a constant offset function of the form $f(x) = Ae^{Bx} + C$ with total RMS error equal to 0.858 MHz. The dissipated power vs. VDD plot has been fitted with the same model function with a total RMS error of approx. 19.315 nW. Exponential fits of this specific form were chosen on the combined basis of the shape of the curves and some trial and error. Finding an analytical model for these behaviours is not trivial if effects such as channel length modulation are to be considered and thus such undertaking lies outside the scope of this thesis.

The simulated values of VDD lie within the realistic window of what one may expect from optical power harvester elements to provide under reasonably powerful illumination. Power dissipation is generally at μW or sub-μW level within this range, whilst frequencies of operation are in the 10s of MHz range. This frequency is theoretically good enough to be able to provide micro-power when used with pF-size capacitors. Supplied power will, of course, be scale with the final voltage (determined by the number of stages in the pump), the size of the capacitors used in the pump and the frequency with which they are allowed to shuttle charge farther down the cascade.

Note: in our charge pump we used capacitors in the vicinity of $50\,pF$ in size. This means that for $0.5-1.5\,V$ unregulated VDD ranges the supplied power will be roughly equivalent (within 50%) with the product $C \cdot f$, where C represents the capacitance at each stage of the charge pump and f the frequency at which it is clocked. Given this information we can already estimate that an upper bound on the power output of the charge pump should theoretically be in the 10s of μW. Under this light, a ring oscillator that consumes $1\,\mu W$ can be potentially seen as power-hungry.

C.2 Drive strength booster

The drive strength booster (DSB) is a one-input-two-output system that receives a signal from the ring oscillator and amplifies it to the point where it can drive the charge pump's complementary clock inputs. Thus, it is important to know how it behaves in terms of power dissipation (related to our power budget), and duty cycle ratio (related to the quality of the output waveform as a clock signal). Formally, the system could be tested in isolation, i.e. receiving an ideal input waveform from a ring oscillator 'simulator' voltage source, however given the simplicity of the system it was deemed more efficient to test the DSB directly with the ring oscillator attached to its input.

The DSB system was tested under various VDD conditions but with no output load. Under the nominal $0.5\,V$ supply and with no external load at the output of the cascade the ring plus DSB complex yields a digital waveform with a frequency of $16.12\,MHz$ at a duty cycle ratio of approximately 49.11%. The total, time-averaged dissipated power measures approximately $112.55\,\mu W$. Note: the ring oscillator plus DSB system is slightly slower than the ring alone, which runs at $16.68\,MHz$. This is a normal consequence of adding a load capacitance within the ring loop. The table of results can be seen in C.2.

Table C.2: Key metrics extracted from the clock generator assembly (ring oscillator plus DSB): output frequency, power dissipation but also duty cycle ratio as a function of supply voltage. The duty cycle ratio represent the amount of time spent by the system output CLK in the 'up' phase.

CLK GEN PERFORMANCE			
VDD	f	P	D
V	MHz	μW	%
0.40	2.80	12.04	49.45
0.45	7.36	42.06	49.55
0.50	16.12	112.55	49.11
0.55	29.76	252.95	48.81
0.60	48.88	499.56	48.63

The results from table C.2 have also been fitted to a function of the form $f(x) = ae^{bx} + c$ with a, b, c

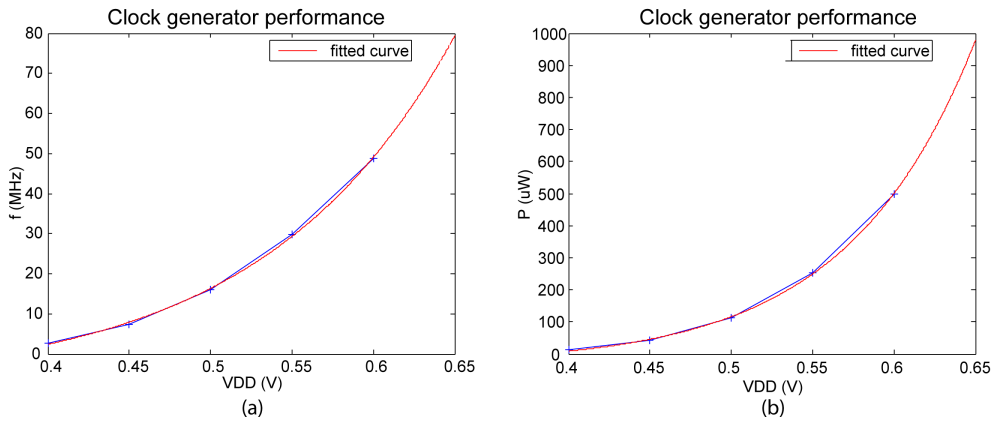

Figure C.2: Data-points and corresponding fits that express clock generator output frequency (a) and energy requirements (b) as a function of applied VDD.

as parameters. The parameter values are summarised in table C.3 whilst the fits can be seen in Figure C.2 plotted along with the original data that gave rise to them. The similar performance variations of the ring oscillator and the clock generator including the DSB are not surprising given that both ring oscillator and DSB are simply cascades of inverters (i.e. consist of the same basic circuit elements) and that the ring oscillator controls the frequency at which the entire chain toggles. As expected, the beat pattern generated by the ring oscillator plays a crucial role in determining the average power consumption of the DSB.

Table C.3: Fit parameters that describe the function of clockgenerator output frequency (f(VDD)) and energy requirements (p(VDD)) as a function of VDD. The parameters correspond to the generic function model $f(x) = ae^{bx} + c$. Fit RMS errors are also given in this table. For f(VDD) the error is measured in MHz and for the p(VDD) function in μW.

FIT PARAMETER VALUES				
Function	a	b	c	RMSE
f(VDD)	0.34	8.54	-7.94	0.6267
p(VDD)	0.25	12.77	-33.02	5.1948

Consideration was given to the matter of testing the system with a simulated load, however given the complicated nature of the load presented to the clock generator system by the charge pump it was decided that the most efficient route would be simulation directly with the charge pump attached to the system. Simulating the load presented by the charge pump by means of a simple RC equivalent would have implied accounting for the pump capacitors and the impedances through which they are connected to other nodes. At nominal $0.5\,V$ arriving from the optical harvesters the resulting clock frequency is equal to $16.1\,MHz$ with a 51.02% duty cycle ratio and a clock generator power dissipation of $286.75\,\mu W$. A quick comparison table summarising the behaviour of the clock generator under different loading conditions is provided in C.4.

We notice that the frequency of the complementary clocks sent to the pumps remains very close to the original $16.12\,MHz$ measured when the clock generator system was tested in isolation. Any discrepancies of this magnitude can be easily explained by small errors in estimating the waveform frequency (done manually from raw data). On the other hand, the power dissipation experienced a dramatic increase

Table C.4: Quick comparison table between the vital statistics of the clock generator system under different conditions of output loading.

CLK GEN OPERATIONAL DATA				
	VDD	f_{clk}	P_{sys}	D
	V	MHz	μW	%
NO LOAD:	0.50	16.12	112.55	49.11
LOAD ON:	0.50	16.10	286.75	51.02

as the final stages of the DSB now have to charge and discharge sizeable capacitances. Nominally, the sum of all capacitors connected to each clock node totals approx. $370\,pF$ and yet because of their connectivity their loading effect on the outputs of the clock generator are different to those of a $370\,pF$ capacitor connected directly to GND. Moreover, the changing impedance linking each pump capacitor to neighbouring nodes means that the loading effect on the charge pump is time-varying.

C.3 Charge pump

The charge pump forms the first part of the path of charge arriving from the optical scavengers on their way to the load circuit. For the tests used to characterise the charge pump we have to use two parameters: the voltage supply provided by the optical harvesters and also the load current drawn from the output of the pump cascade. In terms of optical power supply we will limit ourselves to testing at the nominal $0.5\,V$ level, instead choosing monitor performance for a few values of load current. A single test was ran at a supply of $0.45\,V$ in order to show how VDD can affect the output of the system.

At the output of the charge pump the voltage should be high enough in order to support a down-regulation circuit. Ripples are unavoidable, but need to be contained to the maximum extent possible. As such, important metrics that need to be considered in connection to the charge pump will be the limit voltage under current load, the ripple fluctuation, the minimum voltage appearing at the pump output during a cycle at steady state and the overall power sent by the harvester into what we can describe as the 'power-signal' path of the charge pump. By 'power-signal' path we mean the path taken by charge from the input terminal of the pump to the output terminal as opposed to any paths that serve to feed the circuits that keep the pump running (clock generator). Thus within the charge pump the voltage levels can be treated as signals even though the charge they are referring to will ultimately provide power to circuits farther downstream.

A transient simulation lasting $150\,\mu s$ was found to be sufficient in order to allow the charge pump to reach steady state and thereby also enable us to extract all vital information. In order to calculate the important metrics that describe the charge pump we will use the following techniques: a) The limit voltage is determined by computing the voltage integral between the timestamps of 130 and 131 microsenconds and then extracting the average value. The system reaches a state very close to equilibrium by that stage. b) The ripple is considered to be the range of voltages at the output node of the pump cascade during a full cycle at steady state. c) The minimum voltage during a cycle at steady state is directly extracted from the waveform. d) The output power will be computed by multiplying the limit voltage with the load current used for the simulation.

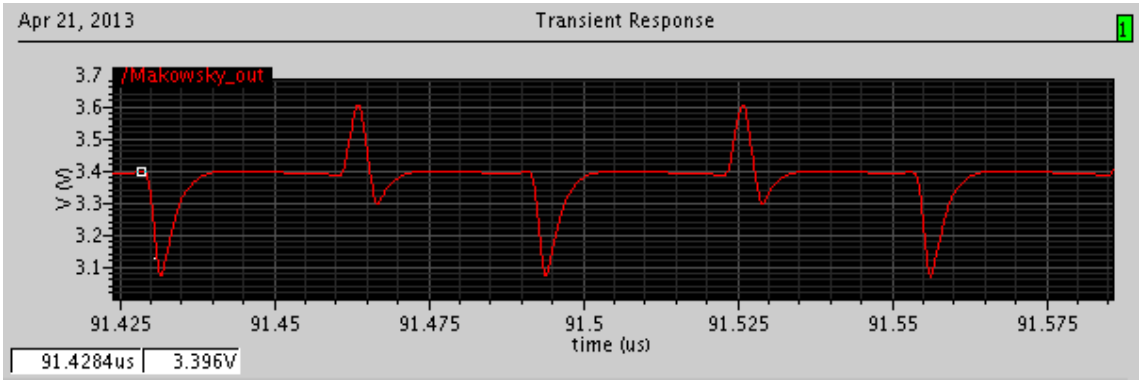

Figure C.3: Typical voltage waveform taken at the output of the charge pump cascade at steady state. The ripples are large in amplitude with the downward leg being most prominent. For relatively long periods of time the voltage stays at a 'plateau' however. Parameters: VDD $= 0.5\,V$. $I_{load} = 12\,\mu A$.

A table of data relating to the charge pump can be seen in C.5. A typical pump output waveform can be seen in Figure C.3. At VDD $= 0.45\,V$ we notice that the asymptotic limit towards which the system output tends is significantly lower than that under a $0.5\,V$ supply with more than double the load current. This shows how crucial the input voltage is for the operation of the charge pump. This fact is also reflected at the power output under the lower $0.45\,V$ supply. Also: the ripple seems to depend on VDD in magnitude. This is expected given the mechanism by which the voltage 'steps' generated by each stage in the charge pump arise.

Table C.5: Key operational data concerning the charge pump module of the PMU. The parameters are separated from the measured results by means of a vertical line. V_{lim}: average voltage towards which the system tends at steady state. P_{use}: useful power flowing into the load of the charge pump. V_{min}: lower bound of the voltage waveform at the pump output at steady state.

CHARGE PUMP DATA					
VDD	I_{load}	V_{lim}	Ripple	P_{use}	V_{min}
V	μA	V	mV	μW	V
0.50	0	3.94	524	0.0	3.65
0.45	12	2.17	480	26.1	1.86
0.50	12	3.38	528	40.8	3.08
0.50	25	2.88	520	72.0	2.55

In order to understand the effects of the asymptotic approach of a system towards steady state we compared maximum voltages (ripple peaks) at different moments in time. In the measurement round under a VDD of $0.45\,V$ the ripple peaks around the 130 microsecond mark rose to $2.33\,V$ whilst at the end of simulation (timestamp of 150 microseconds) ripple peaks were reaching $2.35\,V$. In the $0.5\,V$ case with a $12\,\mu A$ load current the equivalent voltages with timestamps of 130 and 150 microseconds were $3.61\,V$ and $3.61\,V$ respectively. The convergence is clearly quicker in the second case, a fact that can be potentially explained by the higher frequency of operation in the second case. With many more clock cycles available per unit time, the system in the second case can reach the number of clock cycles needed to allow the system to settle at equilibrium (or very close to) much more quickly, even though its limit voltage will be higher. Determining the exact interplay of these two factors lies outside the scope of this thesis.

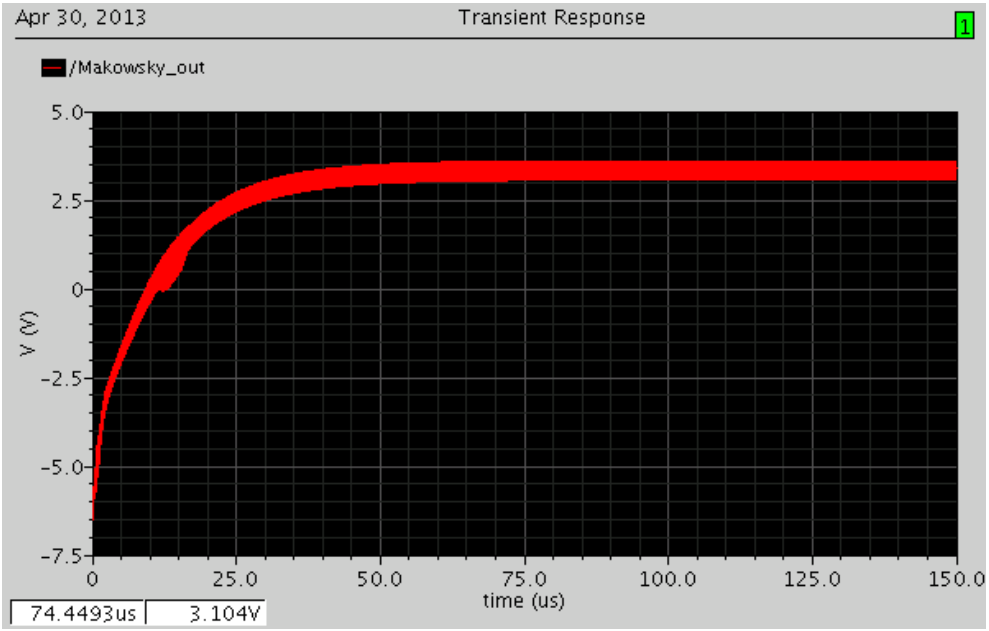

Figure C.4: Voltage waveform taken at the output of the charge pump cascade over an entire test run. Parameters: VDD $= 0.5\,V$. $I_{load} = 12\,\mu A$.

In order to consider these observations as valid indications that the system is still converging it is not necessary to assume that the peaks follow the development of the average voltage over a cycle over time. At equilibrium, the pump output voltage waveform should become perfectly periodic, so any change in peak voltage per cycle indicates that steady state hasn't yet been reached. However, because it is the simple measure of voltage ripple peaks that we are using for our comparison it is not possible to tell whether the relatively small change in ripple peak values over the course of 20 microseconds is reflected in a similarly small change in average output voltage. Nevertheless, visual inspection of the resulting output voltage curve indicates that assuming this is the case would not be entirely misguided. Figure C.4 displays the evolution of the pump output node voltage waveform over the entire duration of a measurement run (VDD $= 0.5\,V$, $I_{tail} = 12\mu A$). The ripples are large in amplitude as evidenced by the thickness of the trace that hides large numbers of high frequency oscillations.

An important point to be made concerning the validity of the test strategy concerns the current demand from the power scavenger over time. Closer inspection of the $0.5\,V$ VDD and $12\,\mu A$ load current case shows that the power drawn from the power harvester element varies wildly over time, showing peaks that reach into the $2-3\,mA$ range but most of the time remaining below $0.5\,mA$. Such large excursions in the current required from the system hint towards the fact that under constant illumination, the voltage across the power harvester elements is likely to experience large excursions itself. Indeed we expect that during those moments when the inverters of the clock generator toggle states and the power harvester elements are faced with with the charge requirements of over 10.000 minimum size inverters, the overall impedance 'seen' by the harvester element will be much lower than in between such moments. Combining this information with the macromodel for our optical power capture element we can infer that the voltage it can provide to the PMU will vary over time. We have already seen how dramatically this affects our ring oscillator and now also the ability of the system as a whole to shuttle charge through the pump cascade.

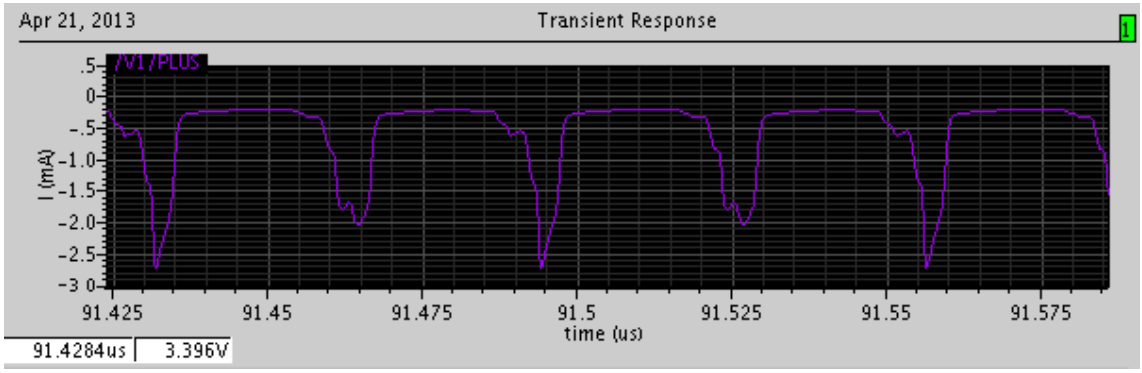

Figure C.5: Typical plot of current drained from the optical power scavenger at steady state. Current sent to the load is not included but at around $12\,\mu A$ remains insignificant. Parameters: VDD $= 0.5\,V$. $I_{load} = 12\,\mu A$.

Note: this effect could have been observed by the point at which the DSB module was tested, but its presentation within the charge pump section places the problem into a much better perspective.

C.4 Voltage reference and start-up circuit

The voltage reference block with its incorporated start-up circuit is tasked with automatically setting its output to a given, stable voltage level once powered up. Therefore, a functional reference will be able to escape the 'all zero' state whereby all circuit nodes (bar VDD) measure $0\,V$ vs GND and no current flows through any branch of the circuit, and reach the desired equilibrium state after a given start-up process. Moreover, the reference should provide very similar voltages for a wide range of Silicon temperatures and strongly reject noise on the power supply node within a wide frequency band. Finally, a well-designed reference will be resistant to process variation and mismatch.

The above aims lead us to identify the following as key metrics of the reference design: a) Start-up circuit success/failure under given power-up protocol. b) Temperature coefficient of the output voltage T_c. c) Power supply rejection ratio as a function of disturbance frequency (PSRR(f)), particularly at those frequencies generated most intensely by the output ripple of the charge pump. d) Performance variation under Monte Carlo simulations (range of output voltages and power dissipation obtained for a given number of Monte Carlo runs). Another important metric that we can add is e) Power dissipation under a variety of VDD voltages (at nominal temperature).

In order to test the start-up procedure we chose to simulate the voltage at the output node of the charge pump as a ramp function starting from the origin ($0\,s$, $0\,V$) and ending at point ($50\,\mu s$, $3.01\,V$). After the 50 microsecond mark the node stays constant at $3\,V$. This corresponds to an increase in the VDD 'seen' by the start-up/reference circuit that is somewhat smoother than the corresponding simulated value given by tests ran on the charge pump with a $0.5\,V$ power input from the power harvesters and $12\,\mu A$ constant load current (see Figure C.4). In order to challenge the system even more another test was ran whereby the start-up ramp is one thousand times gentler than in the previous case. The ramp reaches its final value only after $50\,ms$. Our system starts up successfully in every case. The results of these simulations can be seen in Figure C.6, from which we observe that once the input voltage reaches an approximate value of $1.8\,V$ the reference output value steeply climbs to its desired equilibrium position.

This is also evidenced by an increase in the bias current flowing into the start-up/reference module.

The fact that the voltage reference reaches a non-zero equilibrium point under a lower available VDD in the 'gentle ramp' case demonstrates that the self-biasing circuitry of the combined start-up/reference module is very likely stable only in a non-zero equilibrium configuration for any VDD different than $0\,V$. Proving this definitively requires further testing and circuit analysis which is outside the scope of this thesis.

A DC analysis was run in order to determine the sensitivity of the voltage reference output with respect to temperature. The sweep was performed under a nominal VDD of $3.01\,V$ between the temperatures of $-75^o\,C$ and $+50^o\,C$. The resulting plots are shown in Figure C.7.

Between the two extremes of temperature we notice that current consumption (and similarly power) over-doubles from $5\,\mu A$ to almost $11\,\mu A$ whilst output voltage drops from approx. $1.5575\,V$ to approx. $1.4003\,V$, yielding an average temperature coefficient of slightly less than $1.26\,mV/grad$, which is not particularly low. Given more available time a better reference was intended to be designed.

When it comes to determining power supply rejection ratio (PSRR), care must be taken to ensure that PSRR values are taken at relevant frequencies and under the correct nominal conditions. In our case, we decided to consider a VDD of $3\,V$ as nominal conditions and take measurements at the following frequency set: {10 kHz, 30 kHz, 100 kHz, 1 MHz, 3 MHz, 10 MHz, 30 MHz and 100 MHz}, a set which includes the clock frequency that drives the pump under nominal conditions. Results are summarised in table C.6.

Table C.6: PSRR as a function of disturbance frequency for the start-up/reference module. Up to the $1-3\,MHz$ range PSRR seems to be stable at around $44-45\,dB$.

REFERENCE PSRR(f)	
f	**PSRR**
MHz	dB
0.01	44.44
0.03	44.44
0.1	44.36
0.3	44.66
1	44.88
3	45.27
10	48.07
30	54.66
100	61.94

PSRR was calculated using the formula $PSRR(f) = 20 \cdot log_{10}\left(\frac{|V_{pert}|}{|V_{ripp}|}\right)$, i.e. $20\times$ the logarithm base 10 of the ratio of VDD perturbation to system output ripple magnitudes. The magnitude of the system output ripples was extracted manually from transient simulations whilst the VDD perturbation causing the ripples was set to a 'small-signal' value of $1\,mV$.

We observe that up to the interval between approx. $1-3\,MHz$ PSRR remains fairly stable at a value of around $44-45\,dB$. In broad terms this can be characterised as far from state of the art performance. Higher frequencies induce an increase in PSRR, including at the clock frequency of the charge pump under nominal conditions. The output capacitor is most likely the reason why this effect manifests itself.

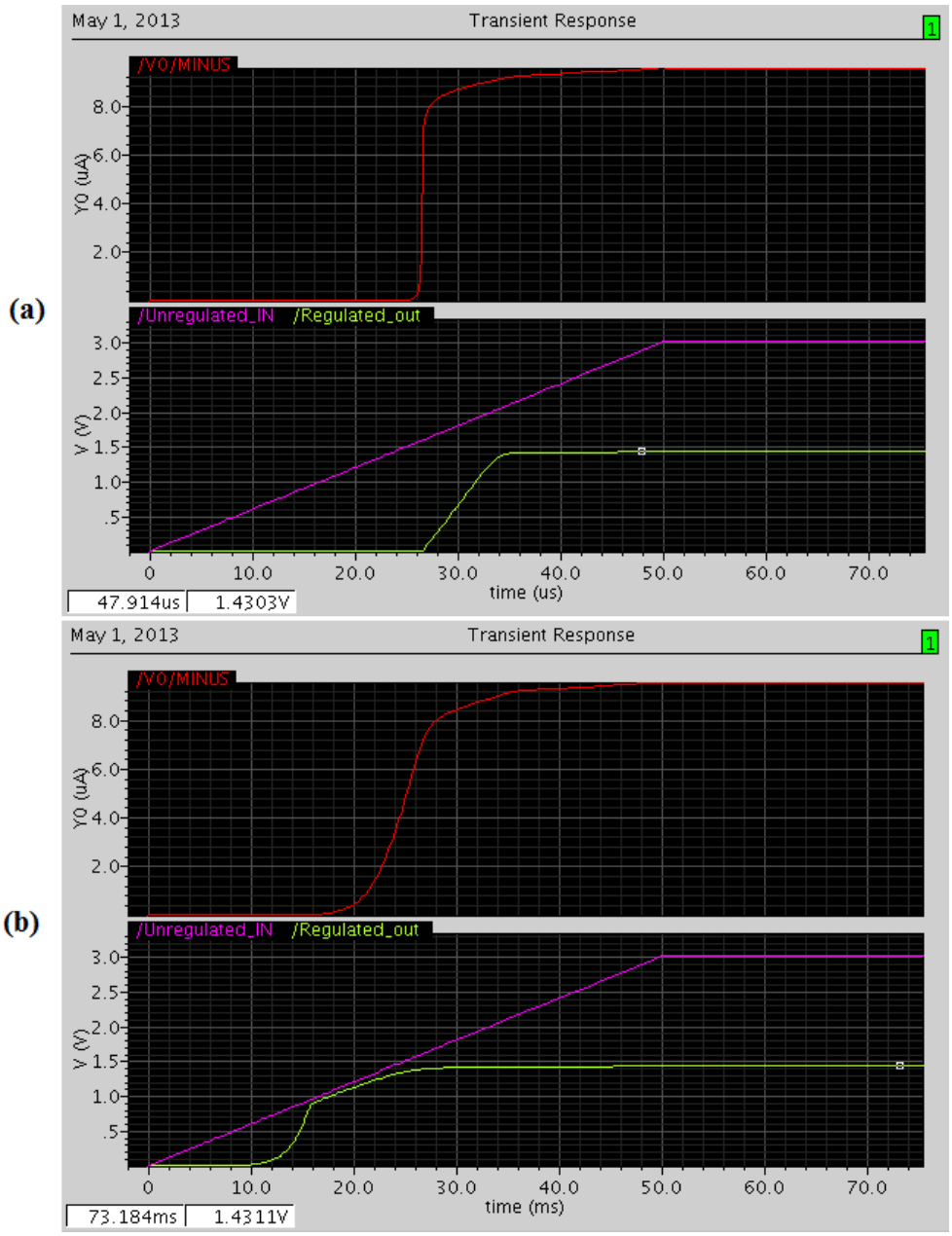

Figure C.6: Typical start-up sequence used for the purposes of testing the ability of the voltage reference circuit to self-bias to the correct level. Panel (a) representsd system behaviour with a 50 μs long start-up time whilst panel (b) corresponds to a 50 ms start-up time. The red trace shows the total current drawn by the start-up/reference module. The pink trace shows the applied input voltage simulating the output of the charge pump (i.e. the VDD of the reference). The green trace shows the output voltage. The timestamp-voltage values at the bottom left of the screen shows the output of the reference at the given point. At steady state the output voltage rises to 1.4311 V in both cases.

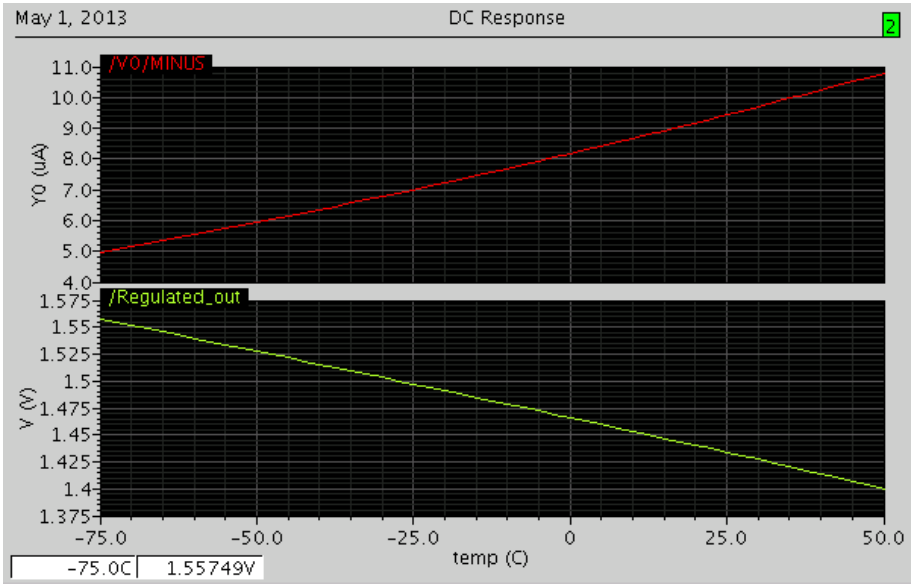

Figure C.7: Variation of voltage reference output voltage and total current consumption with temperature. Red trace: total current consumption. Green trace: Reference output voltage.

In order to further understand PSRR we ran a DC sweep of VDD between $2.5\,V$ and $5.0\,V$. This undertaking should reveal PSRR at DC for a wide range of VDD values. Results are shown in Figure C.8.

The average variation of output vs. VDD voltage given by finding the slope between the extreme ends of the graph in Figure C.8 yields a value of approx. $5.327\,mV/V$. The corresponding dB value is $45.47\,dB$. The overall shape of the V_{out} vs V_{in} curve is visibly slightly convex up with the effect more pronounced at the lower end of the x-axis. This simply demonstrates that the PSRR does not seem to change dramatically at DC between different levels of VDD. Whether this still holds for other frequencies needs to be found out through more measurements. Such detailed analysis of the voltage reference block lies outside the scope of this thesis.

Monte Carlo simulations were run under both process and mismatch variation regimes in Cadence (taking each into consideration in turn) and yielded results that are summarised in Figure C.9. A total of one hundred runs were performed for each type of inconsistency factor set.

It becomes evident from Monte Carlo simulations that the reference block as a whole can suffer serious voltage offsets due to process variation and slightly less severe effects due to mismatch. One possible explanation for this predicament is the reasonably large size of devices used throughout the design of the reference. No transistor is shorter than 1 micron or narrower than 5 microns. Similarly the resistive element used in the beta-multiplier reference measures $15 \times 1.5\mu m$. These are significantly above the minimum feature sizes offered by the technology. On the one hand, the use of such large devices should limit mismatch variation even without the employment of any special layout techniques designed to reduce it. At the same time, process variation is independent of design geometry and will therefore not be ameliorated by our geometrical design choices. This leaves the process variation as the stronger effect.

In terms of power dissipation, mismatch variation is the most damaging, particularly at lower VDD

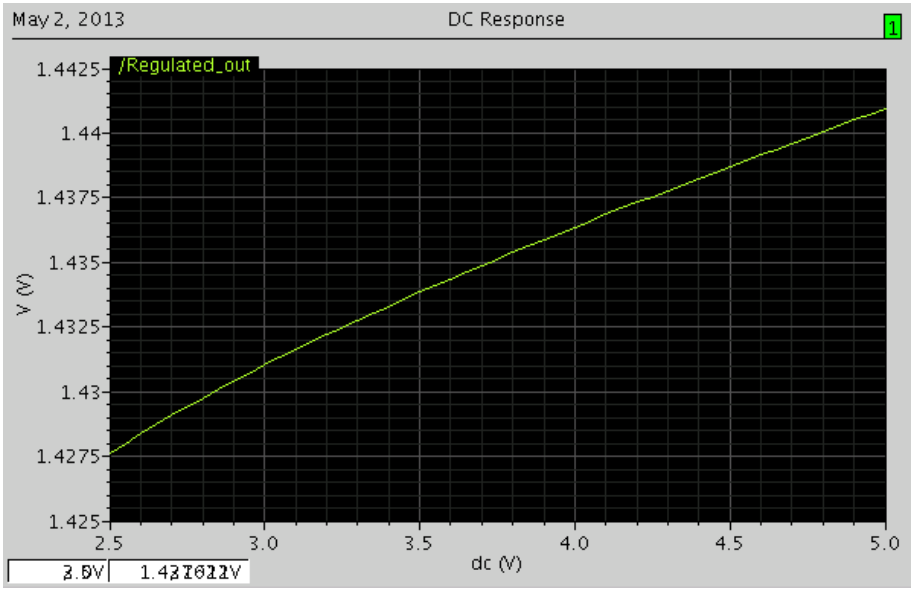

Figure C.8: VDD sweep in the range of $2.5 - 5\,V$ showing reference output variation.

values. There is no obvious reason as to why the amounts of variation induced by these separate mechanisms should have the quantitative relation they have. The fact that the semiconductor property-related variations in the process variation case become exacerbated with temperature while the geometry issues arising by mismatch are not, is also clearly visible in Figure C.9.

Finally, we wish to determine how power dissipation changes as a function of VDD. The analysis leads to Figure C.10. This function is also slightly concave up, which implies that as VDD increases the dissipation of power increases supra-linearly.

C.5 Regulator output stage

The output of the regulator system consists of a simple Op-Amp with Miller compensation and a nulling resistor that controls the gate of a wide output transistor. Key metrics that were extracted out of the system as a whole were: a) GBW value (or to be more accurate unity gain frequency), b) Phase margin, c) Power consumption and d) PSRR as a function of frequency.

Tests were ran with a $10\,pF$ capacitive load and a constant current drain of $20\,\mu A$ attached to the output of the module with the capacitance considered representative and the current draw pessimistic. The bias current used by the amplifier for these tests was nominally equal to $8.4\,\mu A$ excluding the current drawn by the diode-connected transistor tasked with creating the reference current used to generate the branch bias currents within the Op-Amp. The respective bias current value was chosen as a reasonable trade-off between constraining the GBW of the module too much and consuming too much power. Finally, VDD was varied between the extremes of $2.5\,V$ and $5\,V$, defined as the reasonable operating range of the module. If VDD is too low it imposes voltage headroom issues and does not allow the regulated output voltage to remain at the value dictated by the voltage reference. Above $5\,V$ oxide breakdown destroys the MOSFET gates.

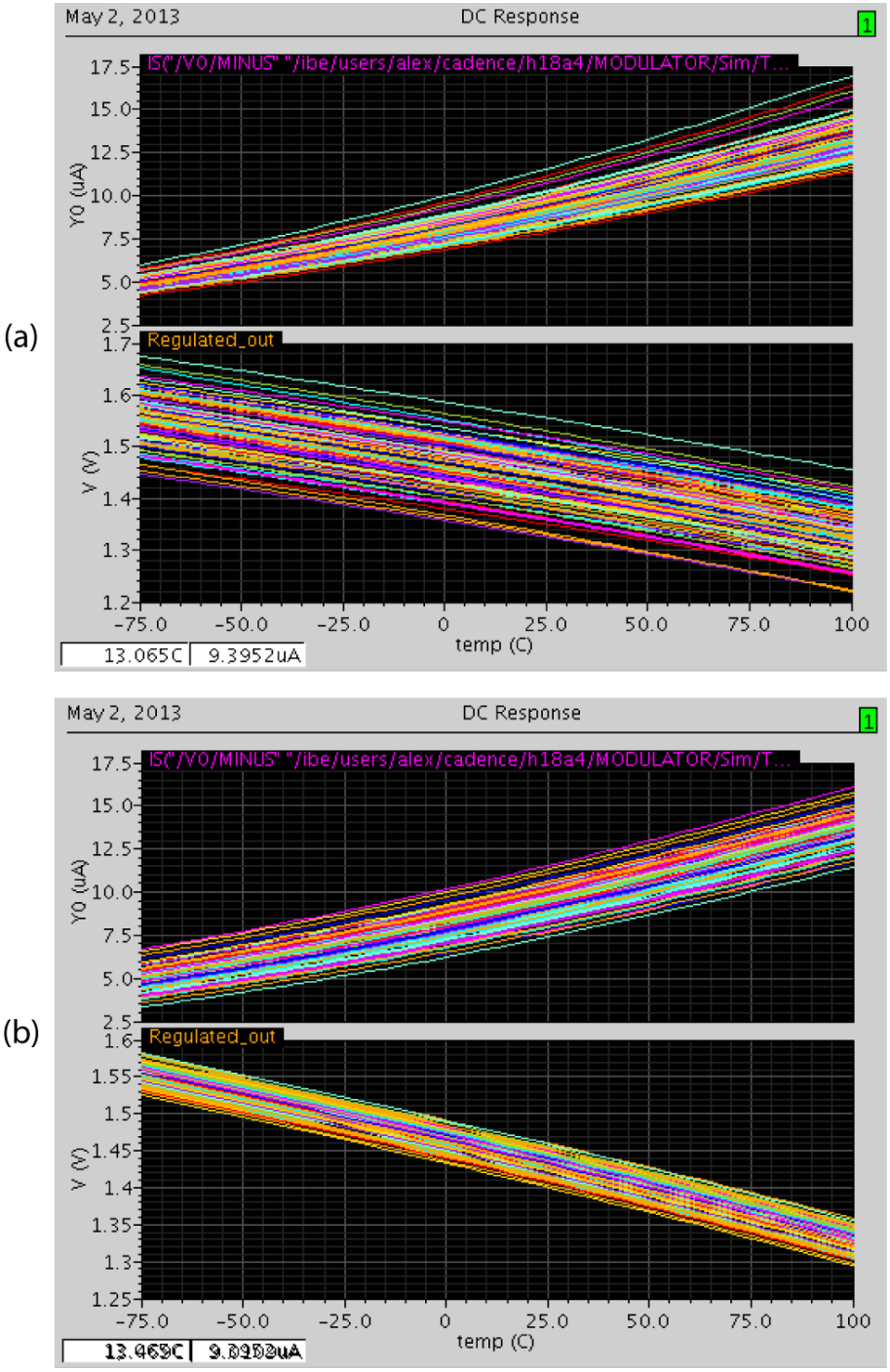

Figure C.9: Process only (a) and mismatch only (b) variation of voltage reference performance under a variety of temperatures. Top trace bundles in each panel represent current consumption whilst bottom traces show output voltage.

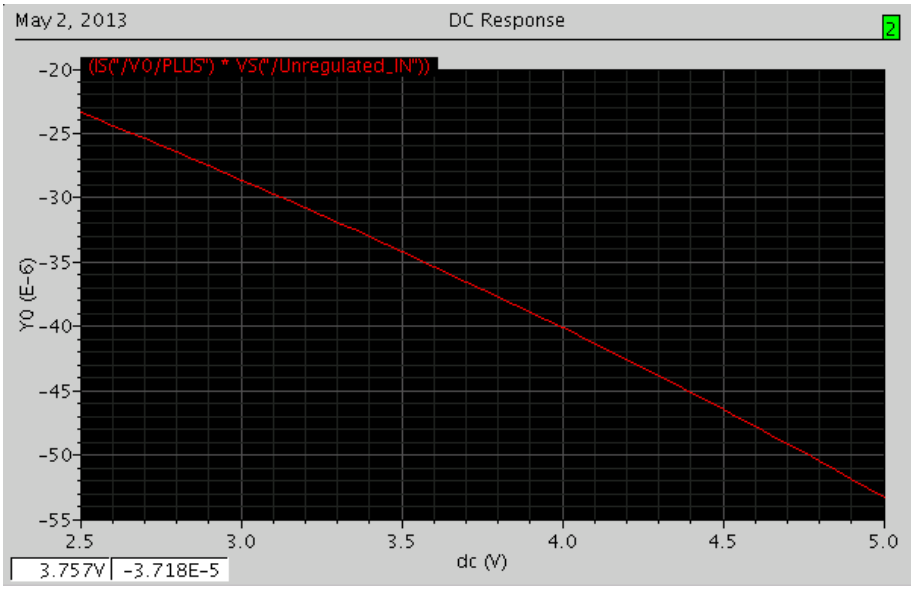

Figure C.10: Power requirement of the start-up/reference block for various values of VDD. The negative sign indicates that power is being dissipated by the block. the y-axis has units of μW.

Note: The bias currents for the Op-Amp branches were generated by use of a small, diode connected transistor from which a specified amount of current is drawn. This 'reference transistor' then shares its gate voltage with its counterparts tasked with creating the bias currents for the branches of the Op-Amp. These latter devices are not immune to channel length modulation. For that reason the actual bias currents tend to deviate from nominal and depend on the exact value of VDD. Moreover, it is important to be aware that these values represent the worst-case scenarios achieved during peak Op-Amp use. During regular use actual power usage is going to be lower by an amount determined by the underlying signal and noise activity throughout the module and its terminals.

Testing on GBW, phase margin and power dissipation resulted in table C.7. GBW and phase margin simulated readings seem to indicate that higher VDD values indeed cause the Op-Amp to be capable of handling higher GBW while at the same time being more stable. Meanwhile, nominal maximum power simply increases according to the simple formula of $P = VDD \cdot I_{nom}$ or power equals power supply voltage times nominal current consumption.

Table C.7: Main metrics describing the regulator output stage. GBW: gain-banwidth product. ϕ_M: phase margin. P: power dissipation.

REGULATOR OUTPUT METRICS			
VDD	**GBW**	ϕ_M	**P**
V	kHz	deg	μW
2.5	142.58	68.82	20.5
3.0	150.54	69.89	24.6
3.5	156.60	70.50	28.7
4.0	162.02	70.94	32.8
4.5	167.11	71.30	36.9
5.0	172.07	71.62	41.0

PSRR values have bee extracted at various frequencies in the $10\,kHz$ to $100\,MHz$. These are shown in table C.8. These PSRR values were obtained by adding a perturbation to the VDD of the Op-Amp only (leaving the output transistor under a separate, stable VDD) and measuring the perturbation at the regulated output node. If we include the perturbation across the output transistor we can expect the PSRR to further deteriorate due to channel length modulation at the output transistor.

Table C.8: PSRR at given frequencies under a VDD of $2.5\,V$.

REG OUT PSRR	
f	**PSRR**
MHz	dB
0.01	44.87
0.03	26.21
0.1	8.31
0.3	1.24
1	7.81
3	16.75
10	27.49
30	36.59
100	44.34

Overall, PSRR seems to show a minimum in the region of $300\,kHz$ where power supply perturbations leak onto the regulated voltage node unabated. Detailed, analytical, small-signal analysis would need to be made in order to determine the exact causes of this phenomenon. This lies outside the scope of the thesis. it must be noted, however, that the poor PSRR of the output stage largely compromises the stabilising effects of the voltage reference, which is why this part of the system would need serious re-thinking.

Appendix D

MATLAB code

In this section lies MATLAB code used for the following purposes:

- Simulating the dynamics dictates by Fick's law of diffusion.
- Simulating the dynamics of ion implantation.
- Simulating combinations of implantation and diffusion.
- Simulating the power delivery to load from a photodiode as a function of photocurrent and load resistance.
- Driving the test bench for characterising our pn structures in power harvesting mode.
- Processing raw I-V traces in order to extract open circuit voltage, short circuit current and maximum provided power.

D.1 Fick's law simulator

```
%%% DIFFUSION SIMULATOR

%%% DECLARATIONS %%%

L = 60; %No of elements of space.
T = 100; %No of time steps.
D = 2; %Diffusion cosntant.

N = zeros(L+1,T); %Matrix holding all diffusion patterns over time.

t = 1; %Time variable.
x = 1; %Space variable.

%%% INITIALISATION %%%

N0 = 1; %Origin set at normalised peak density 1. Rest stays 0.
N(1,:) = N0; %Origin boundary condition.
```

```
%%% RUN PROGRAM %%%
for(t=1:1:T)
    for(x=1:1:L)
        N(x+1,t) = N0*erfc(x/(2*sqrt(D*t)));
    end
end

%%% PLOT RESULT %%%

q = 1:1:L;

subplot(1,2,1)
surf(N);

subplot(1,2,2)
plot(q-1,N(q,10));
xlabel('Location (arb. units)');
ylabel('Doping concentration (normalised)');

hold on;

subplot(1,2,2)
plot(q-1,N(q,25),'r');
xlabel('Location (arb. units)');
ylabel('Doping concentration (normalised)');

subplot(1,2,2)
plot(q-1,N(q,75),'g');
xlabel('Location (arb. units)');
ylabel('Doping concentration (normalised)');
```

D.2 Ion implantation simulator

```
%%% IMPLANTATION SIMULATOR %%%

%%% DECLARATIONS %%%

X = 249; %Elements in the x direction (lateral displacement).
Y = 249; %Elements in the y direction (depth).

A = ceil(X/2); %Mask opening = 1/2 the material length.
B = ceil(X/2); %Centre of material.
Z = ceil(A/2); %Mask 1/2-opening.

N = zeros(X,Y); %Matrix of locations in 2D Silicon sheet.
M = zeros(X,Y); %2nd example matrix.

%Parameters of the well

Sx = 50; %Sigma of distribution in longitudinal direction.
Sy = 20; %Lateral straggle sigma.
m = 150; %Implantation depth.

Sx1 = 50; %Same variables for 2nd example
Sy1 = 5;
m1 = 150;

% Helper variables.

a = 0; %Counter X-direction.
b = 0; %Counter y direction.

%%% RUN CALCULATIONS %%%

for(a = 1:1:X) %Well calculation
    for(b = 1:1:Y)
```

```
        N(a,b) = 0.5*exp(-1*((b-m)^2)/(Sx^2))*(erfc((a-Z-B)/(sqrt(2)*Sy))...
            - erfc((a+Z-B)/(sqrt(2)*Sy)));
        M(a,b) = 0.5*exp(-1*((b-m1)^2)/(Sx1^2))*(erfc((a-Z-B)/(sqrt(2)*Sy1))...
            - erfc((a+Z-B)/(sqrt(2)*Sy1)));
    end
end

%Artificially add mask.

for(a = 1:1:X)
    for(b = 1:1:5)
        N(a,b) = 0.5;
        M(a,b) = 0.5;
    end
end
for(a = B-Z:1:B+Z)
    for(b = 1:1:5)
        N(a,b) = 0;
        M(a,b) = 0;
    end
end

%%% PLOTTING %%%

subplot(2,2,1)
surf(N,'EdgeColor','none','LineStyle','none','FaceLighting','phong');
ylabel('Lateral location (arb. units)');
xlabel('Depth (arb. units)');
zlabel('Normalised doping concentration');

subplot(2,2,2)
contour(N);
ylabel('Lateral location (arb. units)');
xlabel('Depth (arb. units)');

subplot(2,2,3)
surf(M,'EdgeColor','none','LineStyle','none','FaceLighting','phong');
ylabel('Lateral location (arb. units)');
xlabel('Depth (arb. units)');
zlabel('Normalised doping concentration');

subplot(2,2,4)
contour(M);
ylabel('Lateral location (arb. units)');
xlabel('Depth (arb. units)');
```

D.3 Implantation plus diffusion simulator

```
%%% IMPLANTATION + DIFFUSION SIMULATOR %%%

%% IMPLANTATION STAGE

%%% DECLARATIONS %%%

X = 249; %Elements in the x direction (lateral displacement).
Y = 249; %Elements in the y direction (depth).

A = ceil(X/2); %Mask opening in terms of material length.
B = ceil(X/2); %Centre of material.
Z = ceil(A/2); %Mask 1/2-opening.

N = zeros(X,Y); %Matrix of locations in 2D Silicon sheet.
M = zeros(X,Y); %2nd example matrix.

%Parameters of the well
```

```
Sx = 25; %Sigma of distribution in longitudinal direction.
Sy = 5; %Lateral straggle sigma.
m = 15; %Implantation depth.

Sx1 = 50; %Same variables for 2nd example
Sy1 = 5;
m1 = 150;

% Helper variables.

a = 0; %Counter X-direction.
b = 0; %Counter y direction.

%%% RUN IMPLANTATION %%%

for(a = 1:1:X) %Well calculation
    for(b = 1:1:Y)
        N(a,b) = 0.5*exp(-1*((b-m)^2)/(Sx^2))*(erfc((a-Z-B)/(sqrt(2)*Sy))...
            - erfc((a+Z-B)/(sqrt(2)*Sy)));
        M(a,b) = 0.5*exp(-1*((b-m1)^2)/(Sx1^2))*(erfc((a-Z-B)/(sqrt(2)*Sy1))...
            - erfc((a+Z-B)/(sqrt(2)*Sy1)));
    end
end


%% DIFFUSION STAGE

%%% DECLARATIONS %%%

D = 10; %Diffusion constant.
DelSize = 120; %Delta function response square size.
timespan = 50; %Time steps

%Del = zeros(DelSize,DelSize,timespan); %Delta response.
Nt = zeros(X,Y,timespan); %Time varying N matrix.
Mt = zeros(X,Y,timespan);

c = 0; %Impulse response counter.
d = 0; %Impulse response counter.

e = 0; %Location ordinate of target atom.
f = 0; %Location abscissa of target atom.

%%% RUN DIFFUSION %%%

% Impulse response function

for(t = [1 15])
    t %Show timestamp
    for(e = 1:1:X)
        for(f = 1:1:Y)
            for(c = e-floor(DelSize/2):1:e+floor(DelSize/2))
               for(d = f-floor(DelSize/2):1:f+floor(DelSize/2))
                   if(c>0 && c<X && d>0 && d<Y)
                       Nt(e,f,t) = Nt(e,f,t) + N(c,d)*(1/(4*pi*D*t))*...
                           exp((-1*((e-c)^2 + (f-d)^2))/(4*D*t));
                       Mt(e,f,t) = Mt(e,f,t) + M(c,d)*(1/(4*pi*D*t))*...
                           exp((-1*((e-c)^2 + (f-d)^2))/(4*D*t));

                       % Del(c,d,t) = (1/(4*pi*D*t))*exp((-1*(c*c + d*d))/(4*D*t));
                   end
               end
            end
        end
    end
end

%% PLOTTER SECTION
```

```
%%% PLOTTING %%%

subplot(2,2,1)
contour(Nt(:,:,1));
ylabel('Lateral location (arb. units)');
xlabel('Depth (arb. units)');

subplot(2,2,2)
contour(Mt(:,:,1));
ylabel('Lateral location (arb. units)');
xlabel('Depth (arb. units)');

subplot(2,2,3)
contour(Nt(:,:,15));
ylabel('Lateral location (arb. units)');
xlabel('Depth (arb. units)');

subplot(2,2,4)
contour(Mt(:,:,15));
ylabel('Lateral location (arb. units)');
xlabel('Depth (arb. units)');
```

D.4 Photodiode plus load simulator

```
%%% Diode characteristic simulator with parasitic series resistance.

%Variable declaration.
N = 250; %Number of photocurrent points to solve.
P = 10000; %Number of voltage points to check.
Q = 100; %Number of load impedance points to solve.

X = zeros(N,Q); %Vector of solutions.
x = 0; %Variable holding solutions.
xbest = 0; %Variable holding best solution
y = 0; %Variable holding closest solution.
L = 0; %Low range edge for photocurrent.
H = 0.001; %High range edge for for photocurrent (1mA).
Ipd = 0; %Photocurrent initialisation.

Is = 10^(-9); %1nA reverse saturation current.
Rs = 0; %Parasitic series resistance (Ohm).
Rmin = 0; %Minimum tested impedance.
Rmax = 400; %Maximum tested impedance.
k = 0; %Counter variable.
l = 0; %Counter variable.
m = 0; %Counter variable.

%Running program section

for(m = 1:1:Q) %Sweep impedance values.
    flag = m %Show iteration under process.

    Rs(m) = (Rmax-Rmin)*m/Q + Rmin; %Compute impedance.

    for(k=1:1:N) %Sweep photocurrent values.
        Ipd(k) = k*(H-L)/N;
        for(l=1:1:P+1)
            x = IVfun((l-1)/P,Ipd(k),Is,Rs(m)); %Evaluate error.
            if(l == 1) %If 1st evaluation.
                xbest = x; %Present evaluation is best evaluation.
                y = (l-1)/P; %Store voltage of best solution.
            end

            if(abs(x) < abs(xbest)) %If current solution better than current best.
                xbest = x; %Reassign best solution.
                y = (l-1)/P; %Store voltage of best solution.
```

```
            end
        end
        X(k,m) = y*y/Rs(m); %Create array of solutions.
    end
end

%Plotting section
subplot(1,2,2)
surf(Rs,Ipd,X,'EdgeColor','none','LineStyle','none','FaceLighting','phong');
xlabel('Load impedance (Ohm)');
ylabel('Photocurrent (A)');
zlabel('Delievered power (W)');
```

Note: IVfun represents the function: $I = I_s \cdot (e^{V/V_T} - 1) + \frac{V}{R_s} - I_{photo}$.

D.5 IV characterisation script

This script communicates with a Keithley SMU and extracts I-V plots.

```
%%%KEITHLEY controller%%%

%% INITIALISATION SECTION

%Create instrument
K = gpib('ni',0,26);

%Open instrument
fopen(K);

%Reset & identify instrument
fprintf(K,'smua.reset()')
fprintf(K,'smua.reset()')

%Configure instrument for reading & driving
fprintf(K,'smua.source.func = smua.OUTPUT_DCVOLTS') %Select operating mode.
fprintf(K,'smua.source.rangev = 5') %Select source range regime.
fprintf(K,'smua.source.levelv = 0') %Set Vout to chosen value.

%Add compliance limits
fprintf(K,'smua.source.limiti = 10e-2') %Current compliance limit set.
fprintf(K,'smua.measure.rangei = 10e-2') %Set i range to similar level as...
%compliance.

%Switch instrument output on.
fprintf(K,'smua.source.output = smua.OUTPUT_ON')


%Create variables for running the MATLAB side of the loop.
k = 0; %Loop control variable.
Kmax = 200; %Essentially the No of steps in the loop.

V = 0; %Voltage output variable.
Vstart = -1; %Initial voltage.
Vrange = 2; %Voltage range over sweep round.

Iout = zeros(Kmax + 1,2); %Create output variable (array).

fname = 'test.xls'; %Name of the file to be generated.

q = 0; %Plotting variable for negative voltages.
p = 0; %Plotting variable for positive voltages.

%% RUN

%Start the loop
```

```
for(k=0:1:Kmax)
    %Run a test & read round
    V = Vstart + k*Vrange/Kmax; %Set voltage output level for each element...
    %of the sweep.
    fprintf(K,'smua.source.levelv = %2.4f',V) %Set voltage at instrument.
    pause(0.1); %Pause to allow settling.
    fprintf(K,'print(smua.measure.i())') %Ask Keithley to read value.
    I(k+1,2) = str2num(fscanf(K)); %Read value for current, convert it from...
    %string to number & store.
    I(k+1,1) = V; %Store voltage at which measurement has been taken.
end

%Plotting section
%Plotter variable sweep
q=1:1:floor((Kmax+1)/2) + 5; %Capture negative part of curve + 5 pts from...
%positive part.
p=floor((Kmax+1)/2) - 5:1:Kmax+1; %Opposite of q.

%Actual plotting
subplot(1,2,1)
plot(I(q,1),I(q,2))
xlabel('Voltage (V)')
ylabel('Current (A)')

subplot(1,2,2)
plot(I(p,1),I(p,2))
xlabel('Voltage (V)')
ylabel('Current (A)')

%Save data in xls file
save(fname,'I');

%% SHUTDOWN

%Close instrument
fclose(K);
```

D.6 IV plot processor

This script extracts basic information, such as open circuit voltage, short circuit current and maximum provided power from a given I-V curve.

```
%%% Maximum power and power fill factor calculation script.

function [voc,ish,p,f] = celpow2(I)

%%% DECLARATION AND INITIALISATION

voc = 0; %Open circuit voltage
ish = 0; %Short circuit current
p = 0; %Maximum power. Initialise at 0 to make sure any power generation
%is taken into account within for loop.
f  = 0; %Power fill factor
A = 0; %Variable holding integral of IV curve for each device.

a = 0; %Counter variable.
k = 0; %Auxilliary variable holding sample 'length'.
i = 1; %Point after first/last sign transition.
j = 1; %Point before first/last sign transition.
%sure countdown proceeds smoothly.

%%% MAIN RUN
```

```
k  = length(I(:,1)); %Find number of I-V measurement pairs taken from device.

for(a = 1:1:k) %Sweep all datapoint in search of `interesting' values.
    if(I(a,1) == 0) %At V = 0 (short circuit voltage)...
        ish = I(a,2); %...assign short circuit-current to variable ish.
    end
    if(I(a,1)*I(a,2) <= p) %If power dissipation at certain data point lower than
        %record-holder.
        p = I(a,1)*I(a,2); %Update record holder.
    end
end

if((I(1,2) + I(k,2))/2 > I(ceil(k/2),2)) %If the IV is concave.
    'Function: Concave'
    for(a = 2:1:k) %Sweep again ignoring 1st element.
        if(sign(I(a,2)) ~= sign(I(a-1,2))) %If sign transition detected.
            i = I(a,1); %Voltage just after transition.
            j = I(a-1,1); %Voltage just before transition.

            %Compute interpolated voltage.
            voc = j + abs((i-j)*I(a-1,2)/(abs(I(a-1,2)) + abs(I(a,2))));
        end
    end
end

if((I(1,2) + I(k,2))/2 < I(ceil(k/2),2)) %If the IV is convex.
    'Function: Convex'
    for(a = k:-1:2) %Sweep again ignoring 1st element.
        if(sign(I(a,2)) ~= sign(I(a-1,2))) %If sign transition detected.
            i = I(a,1); %Voltage just after transition.
            j = I(a-1,1); %Voltage just before transition.

            %Compute interpolated voltage.
            voc = j + abs((i-j)*I(a-1,2)/(abs(I(a-1,2)) + abs(I(a,2))));
        end
    end
end

f = p/(voc*ish); %Calculate power fill factor.

%Change units:
ish = ish * 1000000; %A to uA.
p = p * -1000000; %W to uW.
f = f * 100; %Plain unitless to percentage.
```

D.7 Doping computer for modulation model

This script was used to create the doping profile used for the worked example in section 4.3.

```
%%%Script for doping in model for modulation.

%%%DECLARATION AND INITIALISATION

%Parameters
l = 1100; %Length of simulated device cross-sectional area in 10x nm.
d = 400; %Depth of simulated device cross-sectional area  in 10x nm.
m = 1000; %Length of square mask in 10x nm.
D = 150; %Maximum implantation depth (depeest and `last' implantation peak).
N = 4; %Number of implantation peaks.
A = 0.75*10^19; %Peak implantation doping for each doping step in (dopants / cm^3).

sx = 40; %Longitudinal spread in 10x nm for each doping step.
sy = 10; %Lateral spread in 10x nm for each doping step.
```

```
%Matrices
M = zeros(N,1); %Holds peak doping depths for each doping process.
sub = zeros(l/2,d); %Substrate matrix holding doping information.

%Auxiliary variables
a = 0; %Sweep variable for doping processes.
b = 0; %Sweep variable that goes through substrate matrix laterally.
c = 0; %Sweep variable that goes through substrate matrix longitudinally.
e = 0; %Sweep through implantation peaks.
q = 0; %Plotter vatiable.

%Flags
DOP = 0; %Doping procedure counter.

%%%MAIN RUN%%%

%Compute peak doping centres.
for(e = 1:1:N)
    M(e,1) = (e-1)*(D/(N-1));
end

%Compute doping profile.
for(a = 1:1:N) %Sweep for each doping process step.
    tic;
    DOP = a %Show doping procedure progress.
    for(b = 1:1:length(sub(:,1))) %Sweep in lateral diection.
        for(c = 1:1:length(sub(1,:))) %Sweep in longitudinal direction.
            sub(b,c) = sub(b,c) + (A/2)*exp(-(c - M(a))^2/(sx)^2)*(erfc((b - m/2)/(sqrt(2)*sy)) - erfc((b + m/
        end
    end
    toc
end

%%%PLOTTING SECTION%%%
M

subplot(1,2,1)
q = 1:1:length(sub(1,:));
semilogy(10*q, sub(1,:))
title('Doping concentration vs. depth at centre of device');
xlabel('Depth in nm');
ylabel('Doping concentration in dopants / cm^3');

subplot(1,2,2)
surf(sub,'EdgeColor', 'none', 'LineStyle', 'none', 'FaceLighting', 'phong')
title('Doping concentration throughout 1/2 cross-section of the device');
xlabel('Depth (nm)');
ylabel('Lateral location (nm)');
zlabel('Doping concentration in dopants / cm^3');
```

Appendix E

PMU circuits

In this part of the appendix we revisit schematics pertaining to the power management unit (PMU) and reproduce them in their full complexity, including all device sizes. Trivial circuits consisting of standard cells provided by the manufacturer, such as the ring oscillator and drive strength booster are omitted.

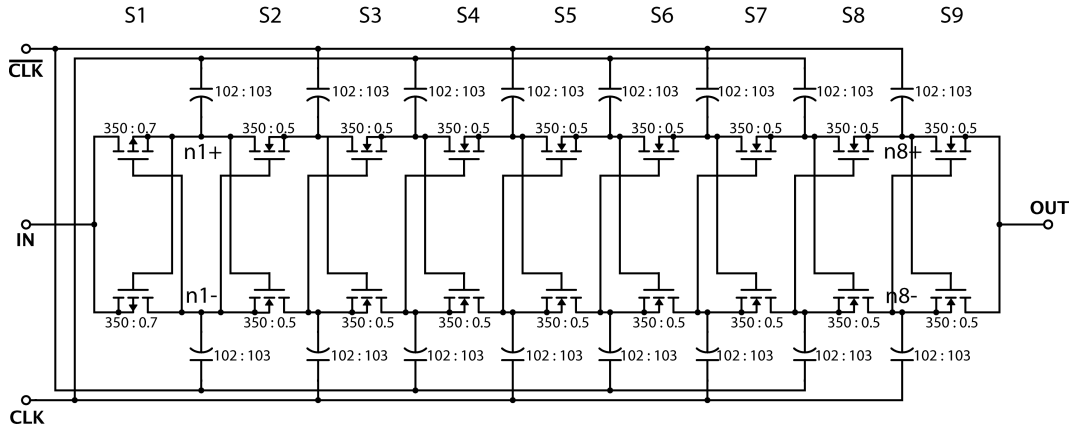

Figure E.1: Charge pump design used in the PMU. The SX indicators, where X is an integer between 1 and 9 indicate the transistor pair stage number. Note the different connectivity pattern for the gates of transistor pair S1. 'nY+' and 'nY-', where Y is an integer between 1 and 8 are node names. Numbers appearing next to each component indicate physical widths and lengths in the format $W : L$ with W denoting width and L length, both in microns.

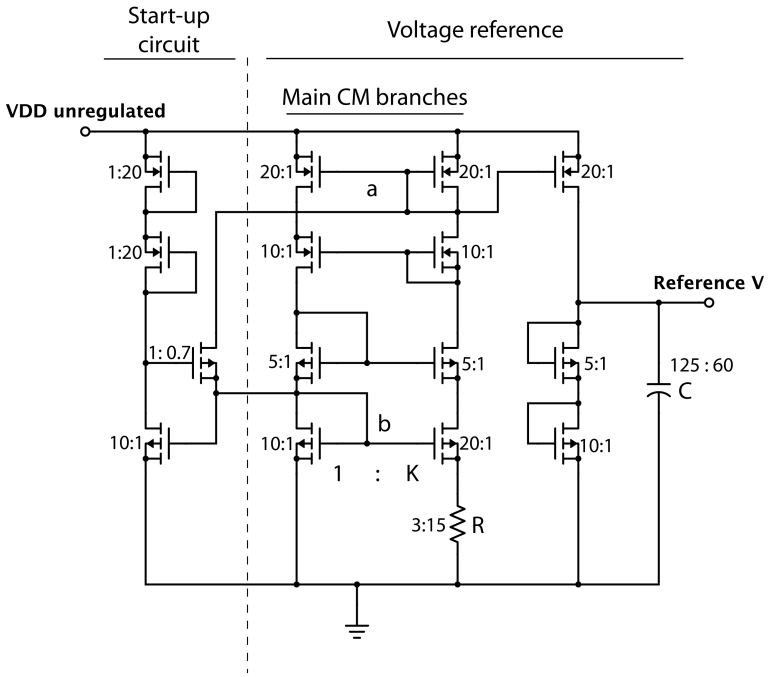

Figure E.2: Voltage reference and associated start-up circuit. The system 'input' consists of the unregulated power supply whilst the output should be a stable voltage. The ΔV_{GS} generating transistor pair is marked by the presence of the '1:K' ratio indicating their effective width ratios. Key nodes have been annotated as 'a' and 'b'. Solid lines above the schematic divide the highly 'laminar' circuit into functional 'strips'. CM: current mirror. Numbers appearing next to each component indicate physical widths and lengths in the format $W : L$ with W denoting width and L length, both in microns.

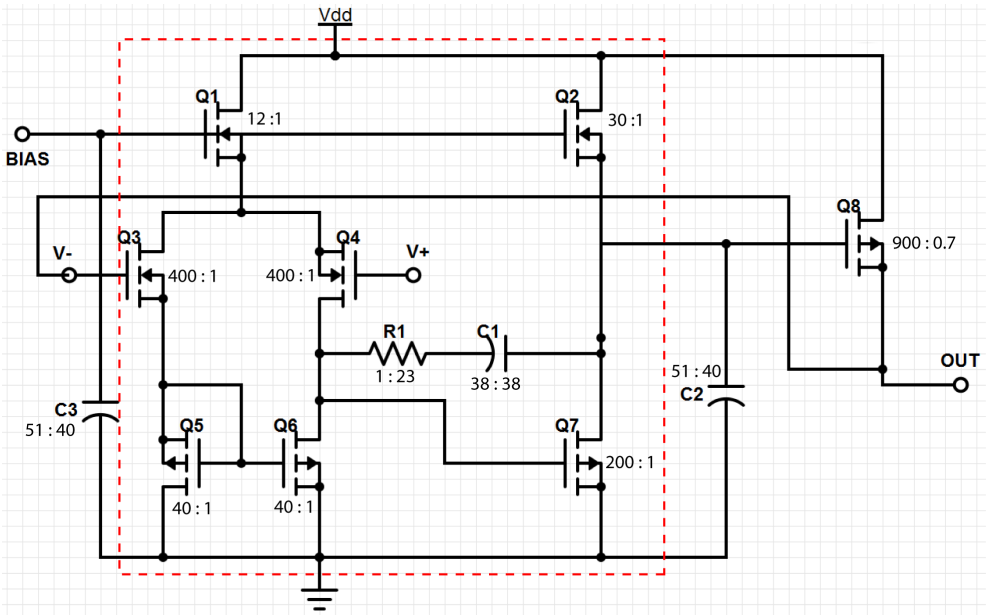

Figure E.3: Output stage of the regulator block. Standard, OpAmp-based design with a wide output transistor. 'Vdd': unregulated power supply. 'BIAS': bias current terminal. 'V+': Plus terminal of the OpAmp - connected to the voltage reference output. 'OUT': regulated output voltage. The amplifier proper has been enclosed into a red, dashed box leaving out the various capacitors and the output transistor. Numbers appearing next to each component indicate physical widths and lengths in the format $W : L$ with W denoting width and L length, both in microns.

Appendix F

PCB diagrams

In this section of the appendix we include PCB diagrams for each design family: Ninja, Svejk and Teddy. In all images within this section, the colour legend is shown in table F.1. A notable, common characteristic in all PCBs is that their centres are all dominated by the footprint of a 64-pin PLCC holder within whose centre two little PTHs (plated through-holes) mark the position where the photodetector is to be soldered.

Table F.1: Colour legend corresponding to PCB diagrams in this section. Top/bottom metal: metal tracks. PTH: Plated through-hole. Silkscreen: Text printed on the PCB. TH: Through-hole.

PCB COLOURMAP	
Colour	**Object type**
Red	Top metal
Blue	Bottom metal
Green	PTH
Turquoise	PCB edge
Gray	Silkscreen
Black	TH

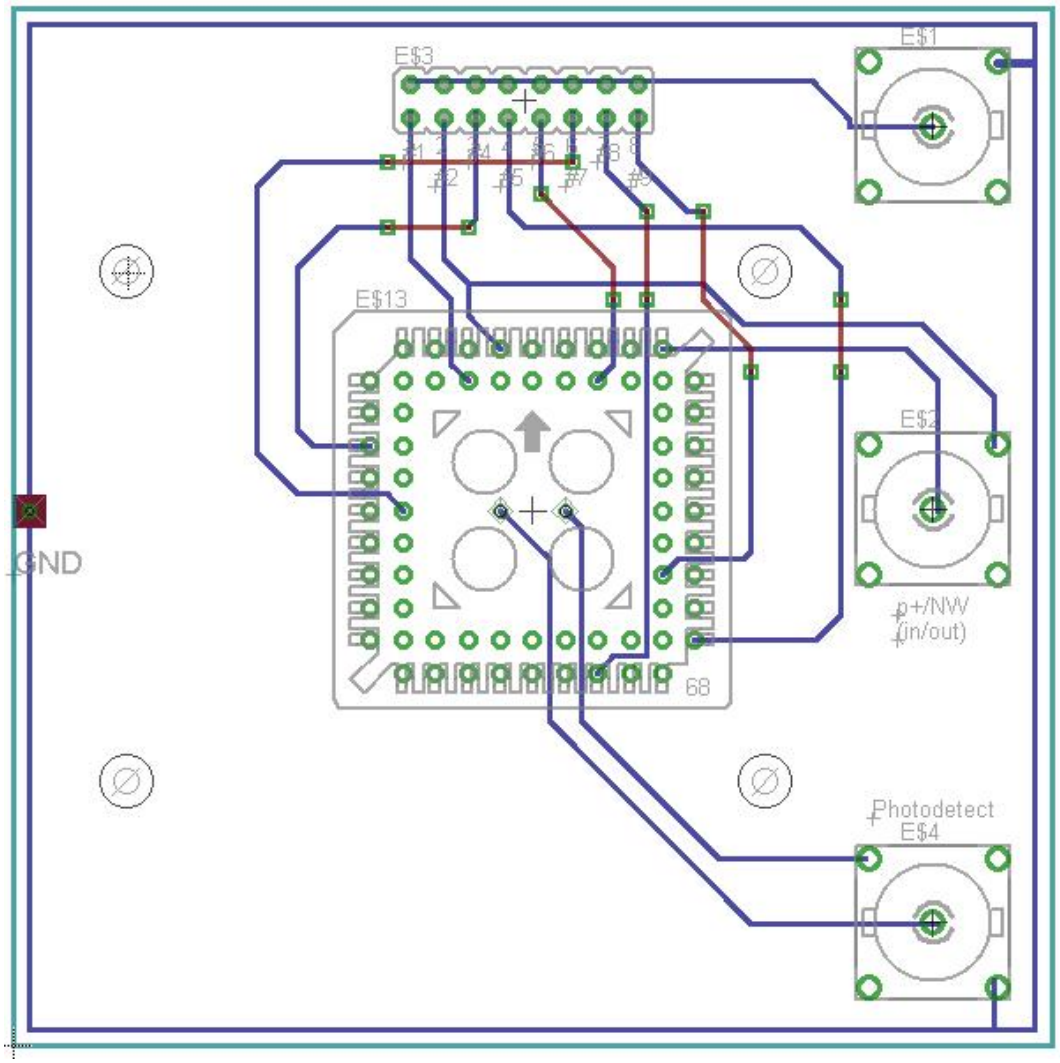

Figure F.1: PCB diagram for the boards that were designed to host dies of the Ninja design family.

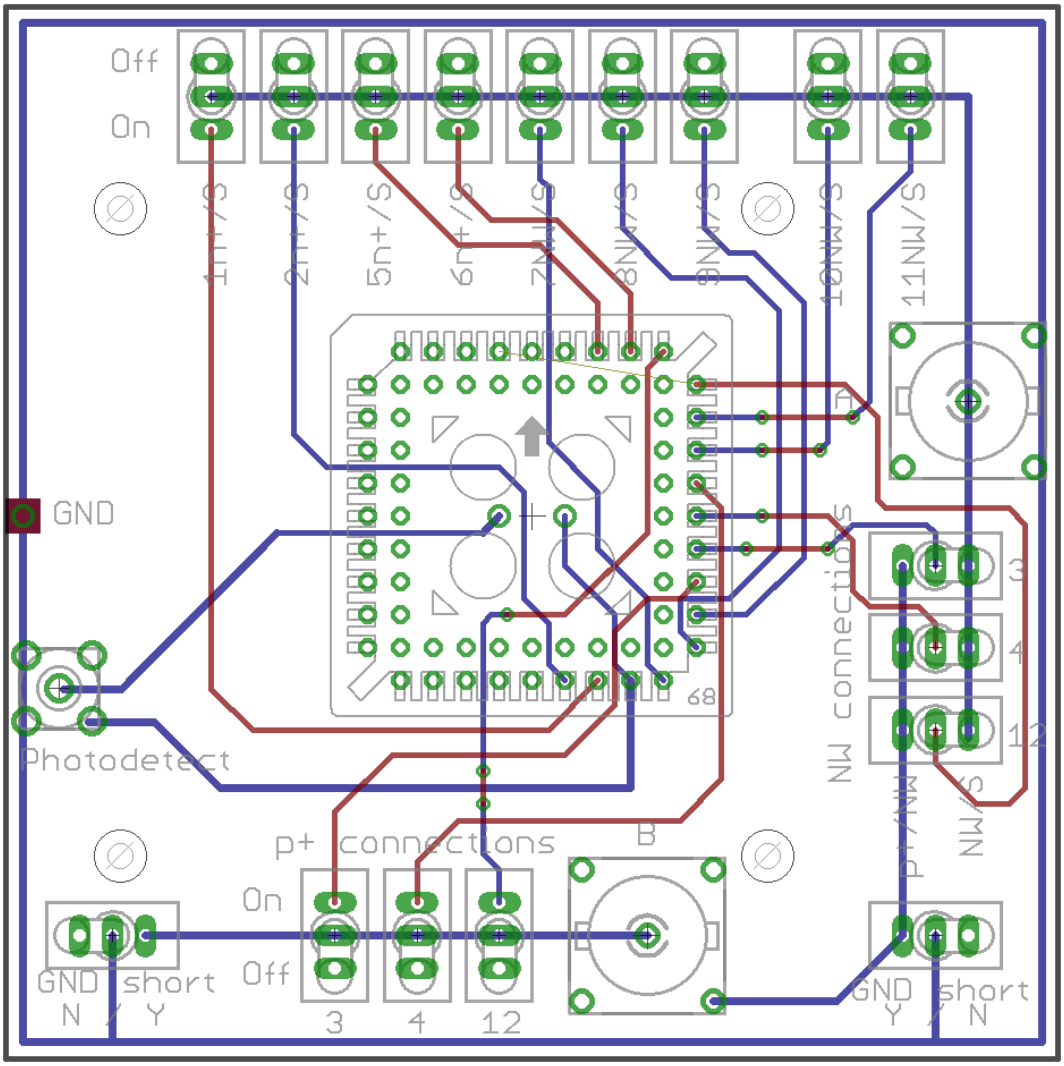

Figure F.2: PCB diagram for the boards that were designed to host dies of the Svejk design family.

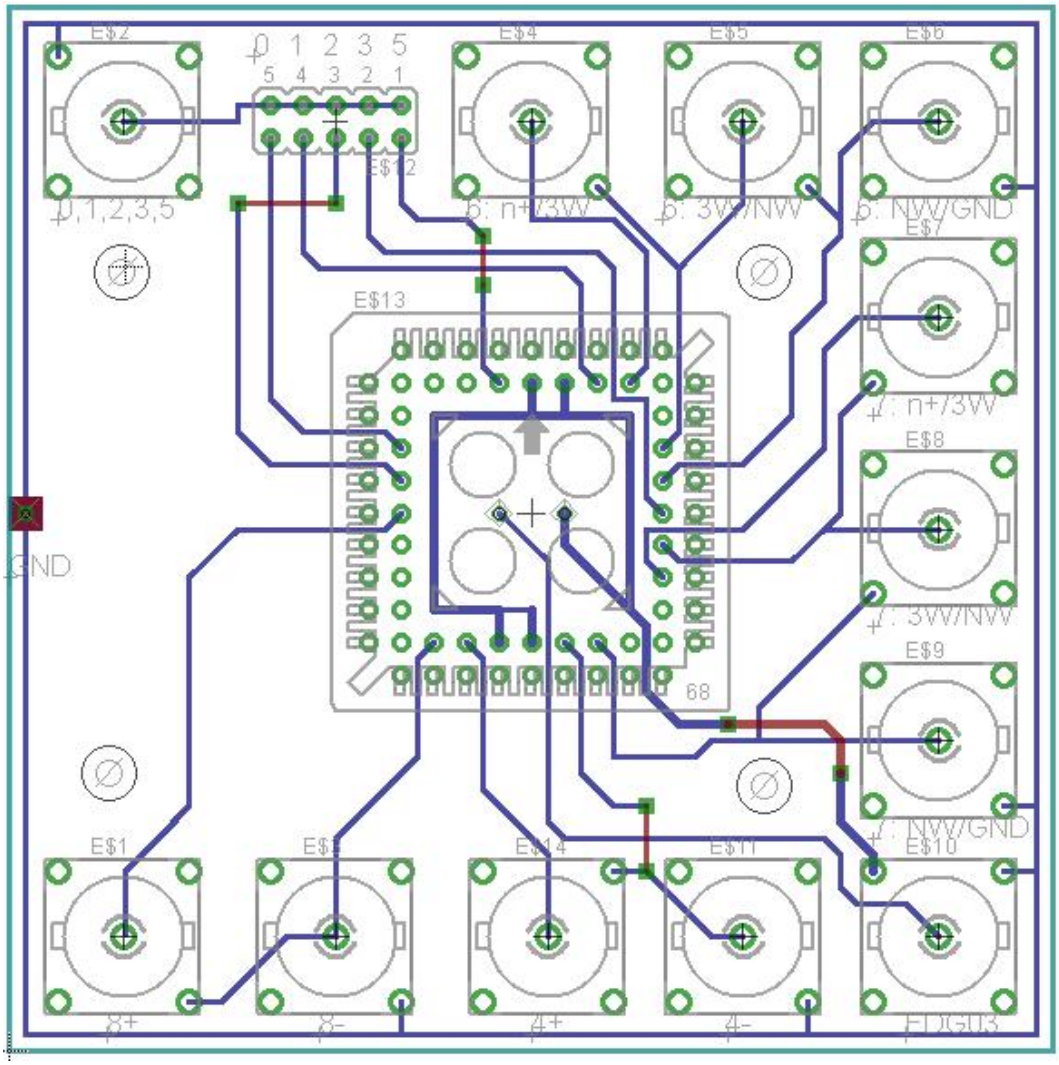

Figure F.3: PCB diagram for the boards that were designed to host dies of the Teddy design family.

Appendix G

Publications

1. A Serb, K Nikolic, TG Constandinou, "Feasibility of an Electro-Optic Link for Bondpad-less CMOS Lab-on-Chips", IEEE Biomedical Circuits and Systems (BioCAS) conference, pp. 353-356, 2011
2. K Nikolic, A Serb, TG Constandinou, "An Optical Modulator in Unmodified, Commercially-Available CMOS Technology", IEEE Photonics Technology Letters, Vol. 23, pp. 1115-1117, 2011.
3. A Serb, K Nikolic, TG Constandinou, "A CMOS-based light modulator for contactless data transfer: theory and concept", Proc. SPIE Photonics West, 2011.