Ultra-low energy photoreceivers for optical interconnects

Ryan Going

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2015-160
http://www.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-160.html

May 31, 2015

Ultra-low energy photoreceivers for optical interconnects

by

Ryan Wayne Going

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ming C Wu, Chair
Professor Eli Yablonovich
Professor Constance Chang-Hasnain
Professor Junqiao Wu

Summer 2015

Ultra-low energy photoreceivers for optical interconnects

Abstract

Ultra-low energy photoreceivers for optical interconnects

by

Ryan Wayne Going

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ming C Wu, Chair

Optical interconnects are increasingly important for our communication and data center systems, and are forecasted to be an essential component of future computers. In order to meet these future demands, optical interconnects must be improved to consume less power than they do today. To do this, both more efficient transmitters and more sensitive receivers must be developed. This work addresses the latter, focusing on device level improvements to tightly couple a low capacitance photodiode with the first stage transistor of the receiver as a single phototransistor device.

First I motivate the need for a coupled phototransistor using a simple circuit model which shows how receiver sensitivity is determined by photodiode capacitance and the length of wire connecting it to the first transistor in a receiver amplifier. Then I describe our use of the unique rapid melt growth technique, which is used to integrate crystalline germanium on silicon photonics substrates without an epitaxial reactor. The resulting material quality is demonstrated with high quality (0.95 A/W, 40+ GHz) germanium photodiodes on silicon waveguides.

Next I describe two germanium phototransistors I have developed. One is a germanium-gated MOSFET on silicon photonics which has up to 18 A/W gate-controlled responsivity at 1550 nm. Simulations show how MOSFET scaling rules can be easily applied to increase both speed and sensitivity. The second is a floating base germanium bipolar phototransistor on silicon photonics with a 15 GHz gain x bandwidth product. The photoBJT also has a clear scaling path, and it is proposed to create a separate gain and absorption region photoBJT to realize the maximum benefit of scaling the BJT without negatively affecting its absorption and photocarrier collection. Following this design a 120 GHz gain x bandwidth photoBJT is simulated. Finally I present a metal-cavity, which can have over 50% quantum efficiency absorption in sub-100 aF germanium photodiodes, which addresses the issue of absorption in photodiodes which have been scaled to near sub-wavelength dimensions.

To Catherine

Contents

List of Figures

List of Tables

Acknowledgments

I first would like to acknowledge the funding that made all of this research possible. E3S (NSF Award 0939514), CIAN (EEC-0812072), Defense Advanced Research Projects Agency (DARPA) E-PHI program under Grant No. HR0011-11-2-0021, Intel, NSF Fellowship (DGE 1106400), the NDSEG Fellowship (32 CFR 168a), and BSAC.

Of course none of this work would have been possible without the input, guidance, and patience of my advisor, Prof. Ming C Wu.

Many technical discussions with Prof. Eli Yablonovich about germanium growth and phototransistors. I also owe gratitude to Prof. Judy Hoyt for informative discussions on germanium growth. I especially am thankful to Erik Chen, who helped me get started in using the rapid melt growth technique that he had developed.

Most of the photodiode work would not have been possible without Tae Joon Seok, who after I left for an internship at Intel, continued the fabrication process we had developed to produce the stellar devices described here. I also am grateful to Myung-Ki Kim, for his helpful guidance with the metal nanocavity design and workings. To Chris Keraly, who it has been a pleasure to work with on the bipolar phototransistor. To Yue Lu for many discussions on the photoreceiver circuit model, and how these devices can tie into a low-power receiver. To the undergraduates I have worked with, Kyle Hsu, Jodi Loo, Tony Duanni, and Darren Lyles who helped characterize many of the devices presented here. To all the members of the group who have helped me get started in the nanolab or with an optical setup, especially Amit, Michael, Karen, Niels, Sangyoon, and Phil.

To Ansheng, Femi, and Ehsan, and many others who made my internship at Intel such a rewarding experience. This work would not be the same without what I had learned while I was there.

The members of E3S, which in many ways was like an extended research group. All of the nanolab staff, especially Sia Parsa, Jeff Clark, Ryan Rivers, and Richelieu Hemphill who worked with me to develop new processes in the lab to process the devices presented here. To Janny, Sherry, and all of the ERSO staff for getting the funds to the right people so that work could continue in the nanolab, and supplies and equipment could be purchased.

To my parents for their love and support, and for culturing my curiosity for science and drive for learning from an early age.

And finally to my loving wife Catherine, who I met at the end of my first year at Berkeley. Marrying you is by far my best accomplishment.

Chapter 1

Introduction

Since the widespread proliferation of the internet, we have become immersed with a wider variety of electronic services, from online banking, to shopping from home, downloading movies directly to your TV, and chatting with friends and family halfway across the globe. We store and retrieve thousands of emails, photos, and songs, and and search entire libraries worth of written materials from our computers or phones. With the advent of the internet of things[5], even more devices, your home appliances and even your car, will be connected to the internet to communicate with each other and the outside world. Running all of these things are massive data centers the size of football fields with power bills larger than those of small cities[6], which contain thousands of individual servers all connected to each other.

Currently datacenters account for roughly 2% of all electricity usage in the United States, and this is projected to increase in the near future[6]. Because the majority of the economic cost and power consumption of a data center is driven by cooling, many are looking to replace power inefficient electrical interconnections with optical interconnects[6].

While electrical wires are remarkably efficient for short reach communications, e.g. we use them daily with USB, ethernet, etc., their efficiency is limited to short lengths. One reason is that the longer the electrical cable, the more energy is dissipated in the wire to communicate. This is because fundamentally the wire is a capacitor which must be charged and discharged to the communication voltage. Its capacitance is linearly correlated to its length, about 200 pF/m[7], so the longer the wire the more energy per bit of data must be dissipated in the cable.

$$Energy/bit = CV^2 \tag{1.1}$$

Optical interconnections, however, do not require charging or discharging of the optical fiber[7], rather there is a fixed cost for the generation of the photons themselves. For this reason, it has generally been easy for optical links to displace electrical links for very long reaches, e.g. under-sea cables, but for shorter reaches, e.g. usb cables, metal lines on the PCB motherboard, electrical lines remain more efficient. In order to eventually displace these shorter electrical links with optics, the light source must be made more efficient, and the optical receiver must be made more sensitive.

As another example, if we look even further out into the future, we can examine the power consumption by metal lines in a modern CPU using numbers from the 2012 ITRS report[8]. Using their metrics for estimations on clock speed, chip area, total metal layers, and their estimates for interconnect power consumption per metal layer, we can estimate the total power consumption for metal interconnects on a CMOS chip. The results show that for 2012, the interconnects were accounting for almost 80% of the total power budget of the CPU. Using their estimates for 2020 with a similar calculation, the interconnect power appears to exceed the total power budget by 91%, which is clearly impossible. What this means is that either major improvements will need to be made to the cooling, such as liquid cooling, the clock speed will need to be reduced, or there will need to be a radical change to the metal interconnects[8]. One potential option is to replace the global metal lines, those running the longest distances across the chip, with optical interconnects.

Table 1.1: ITRS predictions of interconnect power consumption for high performance CPUs

	2012	2020
Max. Power Budget	161 W	130 W
Chip Area	140 mm^2	111 mm^2
Metal Layers	12	14
Clock Speed	3.74 GHz	5.33 GHz
Interconnect Power Index [W/Ghz-cm^2]	1.8-2	2.5-3
Estimated Interconnect Power Consumption	125 W	248 W
Estimated Interconnect Power Consumption as fraction of total power budge	78%	191%

While much emphasis has been placed on making lasers for optical communications more efficient and lower power [9] ultimately it is the sensitivity of the receiver that will determine the total link power consumption since this will determine the optical power level used for communication. In this dissertation, the focus is on exploring what is needed to achieve a very sensitive receiver, and ways to potentially achieve this by demonstrating novel phototransistor devices.

Chapter 2 examines a simple photoreceiver model from an energy perspective to motivate the need to scale down the capacitance of the photodiode and to integrate it with the receiver to form a phototransistor. Since all of the devices in this work are built on silicon photonics, this chapter will also briefly introduce silicon photonics and why it is an essential platform as a solution to the short-scale optical interconnect problem.

Chapter 3 introduces the rapid melt growth technique for obtaining crystalline germanium on oxide, which we have made use of for all of the fabricated devices described in this work. The material quality is characterized and it is shown how we integrated this technique into our device processing. As a starting point, a demonstration germanium pin photodiode

is demonstrated, showing both DC and RF characteristics, which can be used as a point of comparison for further phototransistor devices.

Chapter 4 describes the photoMOSFET, which was largely adapted from the idea and work proposed by the Saraswat group at Stanford, which used a poly-crystalline germanium gate on bulk silicon[10]. Our device is a silicon MOSFET on SOI with a single-crystal germanium gate, which forms the optically active region. The basic device design, along with measurements of a preliminary photoMOSFET are presented. Some thoughts on the future scaling possibility of the device are given.

Chapter 5 describes a bipolar phototransistor which is built as a germanium npn homojunction BJT. The basic design and simulations are presented along with initial experimental results of a floating base bipolar phototransistor. Also some initial simulations are shown which provide a path towards scaling the photoBJT to large gain bandwidth products.

Chapter 6 explores how to scale the photodiode or phototransistor to very small scales. Specifically structures are simulated where the active semiconductor material is on the order of the diffraction limited spot size of the light it is trying to detect. By placing the semiconductor in a well designed optical cavity, resonantly enhanced photodetector can be achieved.

Chapter 7 ties together all of these topics and concludes by proposing how they could be combined and scaled to enable ultra-sensitive photoreceivers on silicon photonics. Some thoughts are also given to potential future directions of this research.

Chapter 2

Low Energy Interconnects on Silicon Photonics

While it is understood that energy consumption by the optical link is a problem which must be addressed in order to make it a compelling solution for replacement of electrical wire interconnects, we must first understand how energy is dissipated in an optical link from a very basic perspective. Several papers have been written describing the constraints for optical interconnects if they are to replace electrical interconnects [7, 9, 11]. Depending on the application and communication distance, they provide different energy targets for optical link performance. In high-performance computing, for example, where interconnects would bridge multiple servers, the energy target for 2020 is put at 1 pJ/bit for the optical link [9]. For 2022, Miller puts a stricter benchmark for 1m or shorter interconnects at 100 fJ/bit [7]. In all of these analyses a similar argument is made for improving the energy per bit consumption by the optical link by means of two primary factors. One is to increase the efficiency of the generation and transmission of the light in the link. The second is to increase the receiver sensitivity by means of reducing the photodiode capacitance[7]. In the following section, we will use a simple photoreceiver circuit model to explain this reasoning, and motivate the remainder of this work in shrinking the photodiode and tightly coupling it to the first transistor stage of the receiver.

2.1 Calculating the Energy per bit for a simplified photoreceiver

We will assume a very simple model for our photoreceiver for future optical interconnects, which is seen in Fig 2.1. The receiver consists simply of a photodiode, which has some capacitance, C_{PD}, that is connected directly to MOSFET transistor in a common-source amplifier configuration. We assume the MOSFET has some gate capacitance, C_g, and that there is some wire connecting the two, which also has capacitance, C_{wire}. The output of the transistor is then applied directly to a following logic stage, which will have some load

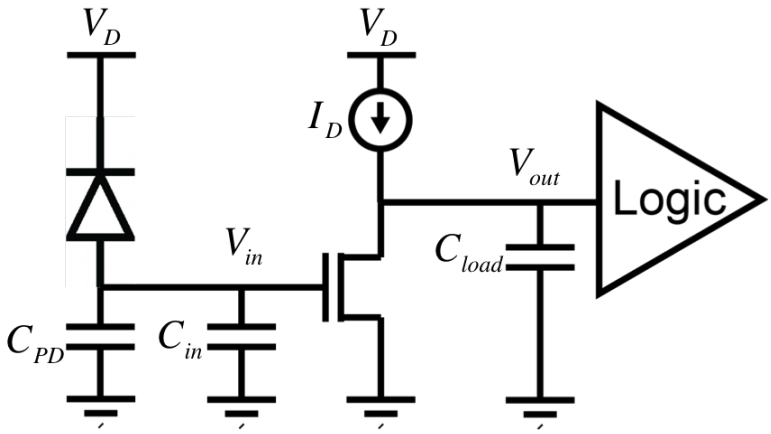

Figure 2.1: Simplified receiver circuit model.

capacitance as well, C_{load}, and which will require a certain voltage, V_{out} to switch to a logical 'on' state. For considering the total link, we will assume our transmission path, or optical fiber, is lossless, and that the light source simply has some quantum efficiency, η_{trans}, associated with its ability to produce photons.

We begin by examining the basic power consumption, which comes from two sources, the transmitter energy to produce the light, and the receiver energy to convert the light back into electrical signal and amplify it, written respectively as the first and second terms in Eq. 2.1. The energy it takes to create the light is simply the photon energy, $\hbar\omega$, with some quantum efficiency, η_{trans}, times the total number of photons per bit, N_{ph}, times the overall bit rate, f_{op}. Not included in this analysis are any transmission losses, though the final expression could incorporate this by simply factoring them into the transmitter quantum efficiency. The power consumption of the receiver is much simpler, as it is simply the voltage and current product going through the receiver transistor. Combining these two together gives the equation for power consumption,

$$P = \hbar\omega\eta_{trans}N_{ph}f_{op} + I_D V_D \tag{2.1}$$

This can be converted into an expression for energy per bit, E/bit, by dividing the total power consumption by the bit rate

$$E/bit = \hbar\omega\eta_{trans}N_{ph} + \frac{I_D V_D}{f_{op}} \tag{2.2}$$

Now to make this a more useful expression, we need to re-write I_D in terms of the input and output voltages and capacitances involved in the circuit, so that we can see the relationship between the photodiode capacitance, the number of photons per bit used to communicate, and the total energy consumption. We can first describe the transconductance gain of the transistor, since we know that a certain amount of gain will be needed to amplify

the input photodiode signal, V_{in} to the needed output voltage, V_{out}. Using a simple expression for g_m,

$$g_m = \frac{I_D}{2V_{ov}} \tag{2.3}$$

where V_{ov} is the overdrive voltage of the transistor, or the amount of voltage applied above the threshold voltage. If we assume to operate the transistor at its peak performance, and that we can ideally make a transistor with a large enough bandwidth to support the gain and operating frequency we desire, then we can imagine out operating frequency will simply be the 3dB point of our amplifier determined by the RC constant of the transistor amplifier and its voltage gain. This can be written in the following way

$$2\pi f_{op} = \frac{g_m}{A_V C_{load}} \tag{2.4}$$

Putting these together we can now write the expression for I_D in terms of operating frequency, voltage gain, and load capacitance

$$I_D = 2V_{ov}2\pi f_{op} A_V C_{load} \tag{2.5}$$

However this is still not the complete story, we must then more accurately tie the transconductance to the amount of amplification, which will simply be

$$A_V = \frac{V_{out}}{V_{in}} \tag{2.6}$$

The final piece is then to express the input photovoltage as a function of photodiode capacitance and the number of input photons, which is the crucial element of this entire derivation. If we imagine the photodiode, input wire, and gate on the MOSFET as a large capacitor, when the photocurrent is generated through absorption of incident light, that photocurrent will charge up that capacitor to a certain voltage, V_{in}. The amount of charge generated is then a relationship between the overall capacitance and the total number of charges generated, which directly corresponds to the total number of incident photons. Writing this more formally we can express it the following way

$$V_{in} = \frac{qN_{ph}\eta_{det}}{C_{PD} + C_g + C_{wire}} \tag{2.7}$$

Combining these we then get the following expression for drain current as a function of

$$I_D = \frac{4\pi V_{ov} V_{out} C_{load} \left(C_{PD} + C_g + C_{wire}\right) f_{op}}{qN_{ph}\eta_{det}} \tag{2.8}$$

Finally we insert this into our original energy per bit expression, to create a more illuminating expression of energy per bit which takes into account capacitance and photon number.

$$E/bit = \hbar\omega\eta_{trans}N_{ph} + \frac{4\pi V_{ov} V_{out} V_D C_{load} \left(C_{PD} + C_g + C_{wire}\right)}{qN_{ph}\eta_{det}} \tag{2.9}$$

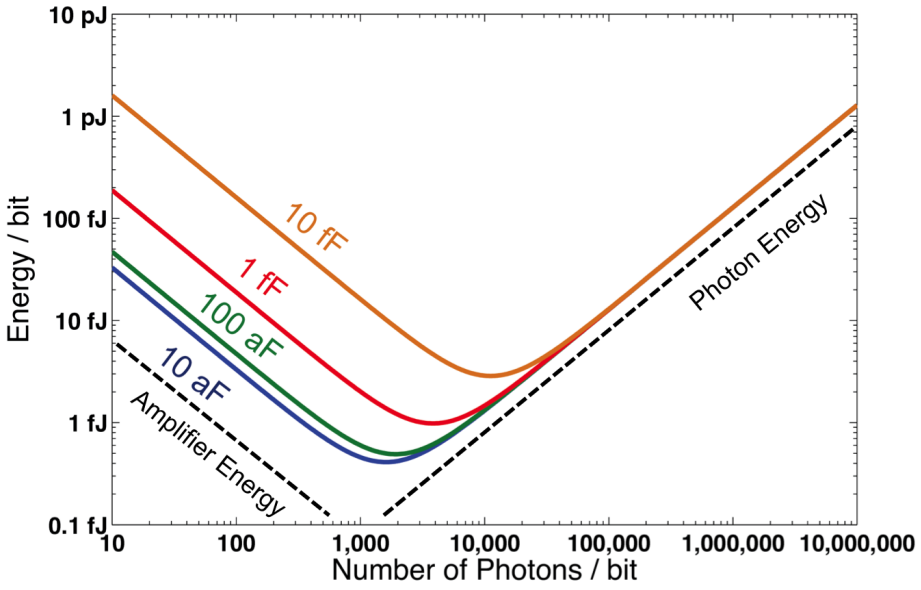

Figure 2.2: Energy per bit relationship with various amounts of photodiode capacitance.

We can now use this equation and put in some example numbers to examine some relationships. For simplicity we will assume that the emitter and detector efficiencies, η_{trans} and η_{det} will be 100%, however this expression is obviously more general. We will then assume that all voltages at 1 V, since this is very close to the line voltage used in modern CMOS[8]. For overdrive voltage, we will assume 100 mV, since it is typical for the line voltage to be roughly only 100 mV above the threshold voltage. Finally, for gate and load capacitance, we will assume 200 aF, since this is fairly typical for current gate thicknesses and dimensions from the ITRS tables[8]. This leaves the remaining variables to be the combined photodiode and wire capacitances, and the number of photons used to communicate. Figure 2.2 plots the energy per bit relationship between these two, showing very clearly the effects of reducing photodiode capacitance.

From the figure, several key trends are evident. First is that in all cases, there is a clear trade-off between using a smaller number of photons and reducing energy consumption, since in all cases there is an energy minimum. This comes about from the two sources of power consumption, transmitter and receiver. Using a small number of photons will reduce energy consumption up to a point because it costs energy to produce the photons in the transmitter. However when the number of photons becomes too few, the photodiode signal is very low, and a large amount of power must be supplied to amplify the signal. Thus for a given amount of photodiode capacitance, there will always be an energetic minimum. Most importantly, as we reduce the capacitance of the photodiode, including the wire connecting it to the receiver, that minimum energy deceases, as does the optimal number of photons per bit. This effectively says decreasing the capacitance of the photodiode increases the overall

receiver sensitivity allowing a more energy efficient optical link.

What this means is the basis of the rest of this work, which is that to create more energy efficient optical links, we must reduce the capacitance of the photodiode, and couple it as tightly as possible to the first stage of the receiver. The closest possible configuration is to monolithically integrate the photodiode with the first transistor to create a phototransistor. This is of course also equivalent to introducing gain into the photodiode itself. In the following section we will explore some recent reports of work following this line of thought.

While the predicted energy per bit and number of photons seems very encouraging, there a number of things this model neglects that will be important in a true optical link. The first assumption of perfect transmitter efficiency is very idealized, and true lasers will certainly have less than ideal efficiency. Additionally, it is common to have an external modulator to the laser as it is very difficult to directly modulate a laser at the higher bit rates typically envisioned for high end optical links.

Second we have totally ignored any signal to noise considerations by focusing entirely on the signal. However this could easily be extended to include those, since the largest noise source in this model is shot noise from the bias current in the amplifying transistor[12]. What this would mean is that overall the minimum energy per bit of all of the curves would likely be shifted upward, since some additional amplification or extra photons would be needed to overcome the noise level for a given bit error rate. Additionally, the curves would also likely feature a more asymptotic curves as the photon number decreased, since the extra amplification needed at low photon numbers would introduce extra noise from the larger bias current, which in turn would require even more amplification to overcome. Thus there would be some minimum photon number below which it would be impossible to achieve a given signal to noise ratio.

Finally we ignore all of the additional circuitry used both on the transmitter side to bias and drive the laser and modulator, and on the receiver side, where additional amplifiers are used to further amplify and modify the signal, such as TIAs, limiting amplifiers, equalization circuits, and clock-data recovery circuits[12]. All of these add additional power consumption, which will dramatically increase the true energy per bit cost of an optical link, which is why it is essential to properly engineer all of the components going into the system. One possible area where there may be a significant gain, is on the receiver side, if the initial photodiode or phototransistor can be made sensitive enough, some of the extra amplifiers or equalizers may be reduced, which would certainly reduce overall power consumption. This however is speculative and beyond the scope of this work.

2.2 Compact integration on silicon photonics

What this model demonstrates is not only that the photodiode must be highly scaled to enable low capacitance, but that the receiving electronics must also be tightly integrated so as to reduce stray capacitance from the wires connecting the two. Because discrete components by definition are not going to be tightly integrated, this calls for a platform where photonic

components and CMOS electronics can be very tightly integrated. One such platform which has been developed very heavily in the last several years is silicon photonics [13, 14], which builds a variety of photonic components on a silicon-on-insulator (SOI) wafer in a CMOS compatible process. While some have used this to attempt monolithic co-fabrication of photonic and electronic components[15], most have simply used the silicon photonics platform as a way to cheaply, and efficiently utilize existing CMOS foundry technology to build high-quality photonic chips that can then be wirebonded[16] or flip-chip bonded[17] to driving CMOS electronics.

For silicon photonics to be an effective, dense, and compact platform, there must first be a variety of components which can replicate the function and performance of their discrete counterparts. The following sections will briefly describe some of these efforts more recently reported in the literature, such as silicon waveguides for transporting and guiding the light, grating and tapered couplers to interface the silicon photonics chip with optical fibers to get light onto and off of the chip, splitters and other optical routing elements, on-chip laser sources, as well as on-chip modulators, and finally detectors, all of which when combined can form a complete optical link.

2.2.1 Waveguides

In creating a compact optical system on a chip with silicon photonics, some of the first work was done to evaluate the use of silicon on oxide as a waveguiding layer for near-infrared light (1300-1600 nm)[14]. Because of the large index contrast between silicon and oxide or silicon and air, light will be confined in the silicon, and since it is transparent at those wavelengths, there should be relatively low losses. Much of the initial work was done on single-mode silicon waveguides[18], and work is continuing to be done to improve propagation losses in waveguides[19], with figures as low as 0.026 dB/cm. Much of the loss is due to either free carrier absorption, which is negated in substrates with extremely pure silicon, and more prominently in scattering losses due to sidewall roughness[14]. As a result, a variety of thicknesses and widths of waveguides have been reported[18], with a more common thickness being 220 nm[14]. Because of sidewall roughness, the bending radius is also a limitation, and much work has been done to optimize sidewalls through photoresist reflow and oxidation smoothing to create very smooth walls, which has been used to demonstrate a silicon ring resonator with Q=2.2 x 10^7[20].

2.2.2 Fiber Coupling

Because the typical cross-sectional dimensions for a single mode silicon waveguide are 220 x 500 nm, but the core diameter for a single mode optical fiber is 10 μm, there is a significant mode overlap problem which must be overcome when trying to couple light onto and off of a silicon photonics chip. There are generally two approaches take, one using a grating coupler[21] and the other involving a tapering of the silicon waveguide to match optical fiber mode[22, 23]. In this work we make use of grating couplers, and an example of a

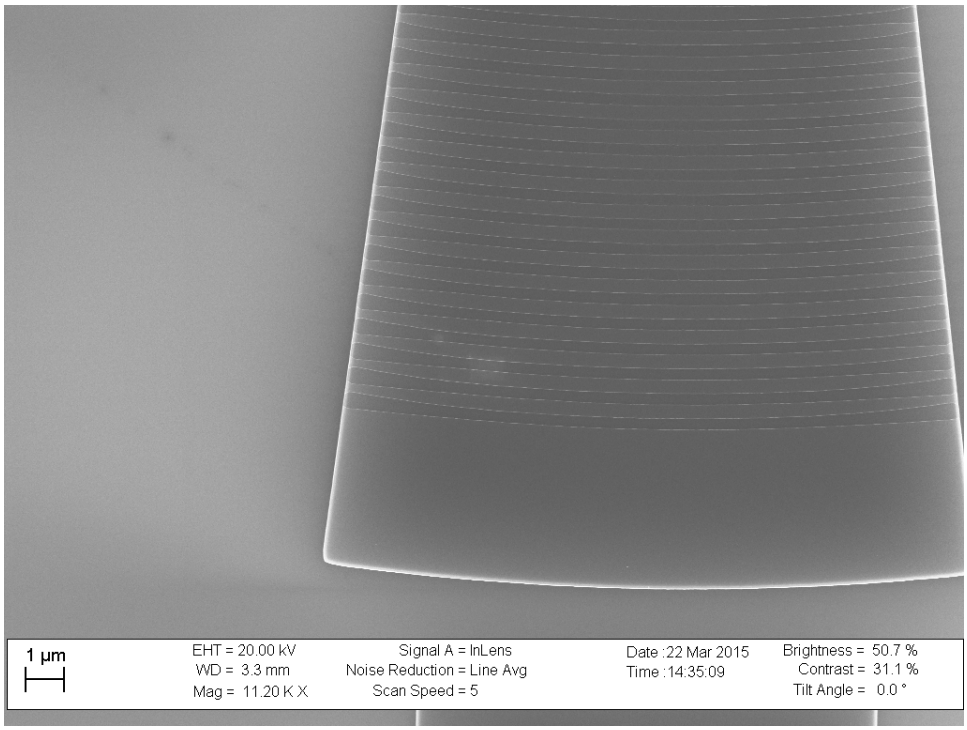

Figure 2.3: Scanning electron micrograph of an example fabricated grating coupler used in experiments in this work.

typical grating coupler can be seen in Fig. 2.3. Grating couplers make use of sub-wavelength diffraction gratings which re-direct light from a normal incident optical fiber to an in-plane waveguide, which is then typically tapered down to the single mode width[21]. While the simplest grating coupler simply uses a shallow etch groove with a 50% duty cycle, and a rectangular lateral shape to match the size of the optical fiber mode[21], this generally results in a coupler which has insertion losses of about 5 dB, with 1 dB spectral bandwidths of 40 nm. To improve on this, many have worked to optimize the burried oxide layer thickness under the grating coupler[24], adjusted the spacing between trenches[25, 26], and added overlay layers[27]. To date some of the best reported insertion losses are around 0.5 dB, but requiring extremely small lithography and exotic substrate preparation[28].

The other method for coupling light onto the silicon chip is through tapering of the silicon waveguide. This can be done by widening the waveguide to more closely match the fiber diameter[22], or utilize a secondary material as a mode-converter with an inverse tapered waveguide, which pushed the light into the lower-index material[23]. While there are benefits over a grating coupler, such as having lower insertion loss, a more broadband coupling, and being polarization insensitive[23], these come at the expense of needing to couple at the edge of the chip, restricting chip layouts, and with reduced tolerances for mis-alignment of the fiber[23]. Typically a tapered or lensed fiber is also used, which has a smaller diameter more

closely matched to a small diameter waveguide[29].

2.2.3 Optical Routing

In order to effectively route the light around the chip, there must also be ways to split the light to different waveguides. This can be done by using y-splitters to split the optical power in half between two waveguides or combine two waveguides[30]. To couple a specific amount of light from one waveguide to another, a coupling length can be specified to create a directional coupler[31]. Many of these can then be used in parallel with multiple waveguides to route light from many waveguide ports to many waveguide ports, such as was recently reported using directional couplers and MEMS actuators[32].

To route light based on its wavelength, such as for wavelength division multiplexing (WDM), a variety of methods have been used. A very simple structure involves using a ring resonator[33], though because of their narrow resonance, they generally must be thermally tuned. Another common wavelength sensitive routing structure is the Echelle grating[34], which is generally quite large, several mm^2 in typical dimensions. Finally there is the arrayed waveguide grating (AWG), which can be compact and generally does not require thermal tuning[35]. Recently AWGs have been demonstrated to be a fully integrated WDM silicon photonic receiver[36].

2.2.4 Light Sources

Although there is continual work done to attempt a silicon laser, because silicon does not have a direct band gap, it generally makes a very poor emitter. There was a demonstration of a silicon Raman laser[37, 38] it still requires an external optical pump. There has also been some initial work to demonstrate lasing in germanium[39, 40] or in germanium-tin alloys[41]. However these are also fairly inefficient at the moment, and require extraordinary doping levels and built-in strain in order to laser at room temperature. For the moment, the preferred method to get light onto silicon photonic chips is through direct wafer bonding of III-V materials[42, 43], which allows for common III-V direct bandgap materials to be monolithically integrated for distributed feedback lasers on silicon waveguides. This is done typically with either an oxide bonding technique or through a polymer adhesive such as BCB. Grating couplers are added to the silicon waveguide to form a DFB cavity for the III-V laser[42].

2.2.5 Modulators

Although silicon does not have a direct electro-optic effect like that of $LiNbO_3$, a commonly used material for discrete modulators[44], it does absorb light in the near infrared through free carrier absorption[44], and large carrier concentrations also change the refractive index. Because of this, silicon modulators are generally p-n diodes, which can be either forward or reverse biased to either inject excess carrier or deplete carriers to shift the refractive index[44].

Because the change is fairly weak, if the modulators are in a Mach-Zehnder interferometer (MZI) configuration, the modulators must be quite long, which incurs both large amounts of optical loss, and large amounts of power dissipation from the electrical bias[44, 45]. Because of this, many are pursuing ring-resonator based silicon modulators, where small changes in refractive index can have a greater effect[44]. Because of their reduced size, ring modulators can also be made to operate very quickly, up to 50 GB/s[46], however like ring resonators for routing, also generally require some thermal tuning. An alternative to silicon is the use of germanium quantum wells as an electro-absorption modulator, utilizing the quantum-confined stark effect[47]. However this type of modulator is less developed and typically has a worse insertion loss than a silicon modulator.

2.2.6 Detectors

Much work has been spent on Ge based detectors[48–57], which will be described in more detail in the following sections. There have also been attempts to induce defect states in silicon for sub-bandgap detection of light by inducing defect states into silicon, but there is generally very low amounts of absorption in silicon in the 1300-1550 nm wavelength range, and so responsivites are generally quite low, and complicated or large structures must be built to incorporate resonance [58]. Others have incorporated tin into germanium photodiodes in order to extend their detection range beyond 1550 nm or to increase the absorption coefficient at 1550 nm [59, 60].

2.3 Previous work integrating germanium photodetectors with gain

From the simple model calculating energy per bit for an optical link, we can see that the photodiode should be connected as closely as possible to the first stage transistor of the receiver electronics. This is to realize the gains of reducing photodiode capacitance by using a short, small-capacitance wire to connect to the receiver. Some have focused on simply packaging the receiver as closely as possible to the photodiode to achieve this, such as with wire-bonding[16]. A method for even closer integration that shows great promise is to flip-chip bond the silicon photonics chip to the CMOS receiver chip[17]. Others have even tried monolithic integration of both photonic and electronic components to increase sensitivity but at a drastically more complicated fabrication process[15].

One of the more traditional ways to incorporate gain into the photodiode is through the use of an avalanche photodiode (APD). APDs work by biasing the photodiode to near the breakdown voltage, where photogenerated electrons and holes will cause impact ionization events as they are collected, creating additional carriers, and thus additional current[61]. APDs generally can have high gain and high speed, and have been used reliably in long-haul telecommunications links[12]. Recently the gain and bandwidth of APDs has been dramatically improved by optimizing material choice, and creating separate regions for absorption

and multiplication with germanium and silicon[4]. While the gain bandwidth product, 340 GHz, is very impressive, the APD requires a voltage of nearly 25 V to achieve this. Several other groups have reported using pure germanium with narrow contact spacings to reduce the voltages down to 6 V or even 3 V[62, 63]. However even these voltages are still well above the standard CMOS line voltage of roughly 1 V. Additionally, avalanche gain is has additional excess noise due to the statistical nature of the multiplication process, which can reduce the sensitivity benefits coming from the integrated gain[12]. Finally, for a fixed voltage, avalanche gain has been shown to dramatically reduce when the ambient temperature is elevated[64]. This is an important consideration for short scale interconnects due to the likelihood of being placed in or near the CPU, which can easily reach temperatures well above room temperature.

The immediate phototransistor that comes to mind springs from the simple circuit shown earlier, where a p-n photodiode is merged with the gate of a MOSFET. We refer to this as a photoMOSFET, and the initial design and simulations were reported by the Saraswat group[10]. However their initial device demonstration did not perform as well as expected, likely due to their use of poly-crystalline germanium, which has excess carrier recombination compared to single-crystal germanium. Because the device relies on storage of holes in the gate material, in this case germanium, if the recombination is too quick, no gain will be seen[10]. We improve on their design by using single crystal germanium and placing the entire device on silicon photonics for better light absorption, which is described in detail in Chapter 4.

Another type of phototransistor design which was pursued by some groups was the junction field-effect transistor (JFET), where a germanium photodiode is grown directly on top of a silicon channel transistor [65–67]. The device functions much like a gated MOSFET, with the channel nominally blocked by the band-bending induced by the Ge-Si interface. However when the germanium absorbs light, holes become trapped, and the channel turns on, leading to increased current and gain. Because the critical feature of the device relies on the germanium/silicon junction, which in general has poor crystal quality, and because the device can only turn off through hole recombination, so far the fastest reported photoJFET has been 8 GHz [67], which had fairly low responsivity. It is very likely the poor quality of the germanium/silicon interface will ultimately limit this particular phototransistor.

Finally, there is the photoBJT, which is in reality simply a bipolar transistor (BJT) where light is absorbed in the base-collector junction[61]. Here the photocariers are injected into the base, much like a traditional bias current being fed into the base. Because the base bias can be applied with light, one can also make a floating base photoBJT where there is not electrical contact on the base. A germanium floating base photoBJT was recently demonstrated by A*STAR[68], who reported a 7 GHz gain-bandwidth product. Through scaling and improvement of the device design, we believe this could be vastly improved to match or even surpass the performance of the best reported APD.

Chapter 3

Germanium Rapid Melt Growth

One of the key challenges for silicon photonics is in the integration of optically active materials onto silicon. Most optically active III-V materials have a dramatically different crystal lattice constant than silicon, which makes direct growth difficult, but in recent years wafer bonding similar to SOI production has occurred to allow for III-V lasers on silicon[42, 69, 70]. However for the detector many had opted towards germanium which can be directly grown on silicon[53] for easier integration and for some devices it can actually have better performance as a detector than its III-V counterparts[4]. It has the additional benefit of being readily CMOS compatible, since SiGe compounds have already been introduced for PMOS channel strain[71] and BiCMOS heterojunction bipolar transistors[72]. There are several growth methods available for obtaining germanium on silicon substrate, but the one we focus on here is the rapid melt growth (RMG) technique[73], which is particularly easy to implement in a university laboratory environment since it does not require epitaxial growth chambers. The process is described here in detail as well as some basic ways to characterize the resulting material. Finally to demonstrate the RMG technique and its resulting material quality, a photodiode is demonstrated on a silicon waveguide.

3.1 Why Germanium

While germanium is an indirect band-gap material, it does have a direct transition at 0.8 eV[74], which corresponds to about 1.55μm wavelength. As a result, it has a much stronger absorption coefficient in the 1.3-1.55μm range than silicon, which is transparent in that range[1]. The absorption coefficient is actually comparable to InGaAsP as well, as can be seen in Fig 3.1. At 1550 nm, the absorption length of germanium is about 8 μm which corresponds to an absorption coefficient of 1250 cm^{-1}.

Additionally, SiGe and Ge compounds have already been introduced into CMOS foundries as a critical element for CMOS transistors and BiCMOS devices. In CMOS, it was introduced to apply strain to the channel in PMOS transistors, beginning with the 45 nm node[71]. And in BiCMOS, it is used to create high-speed SiGe heterjunction bipolar junction transistors

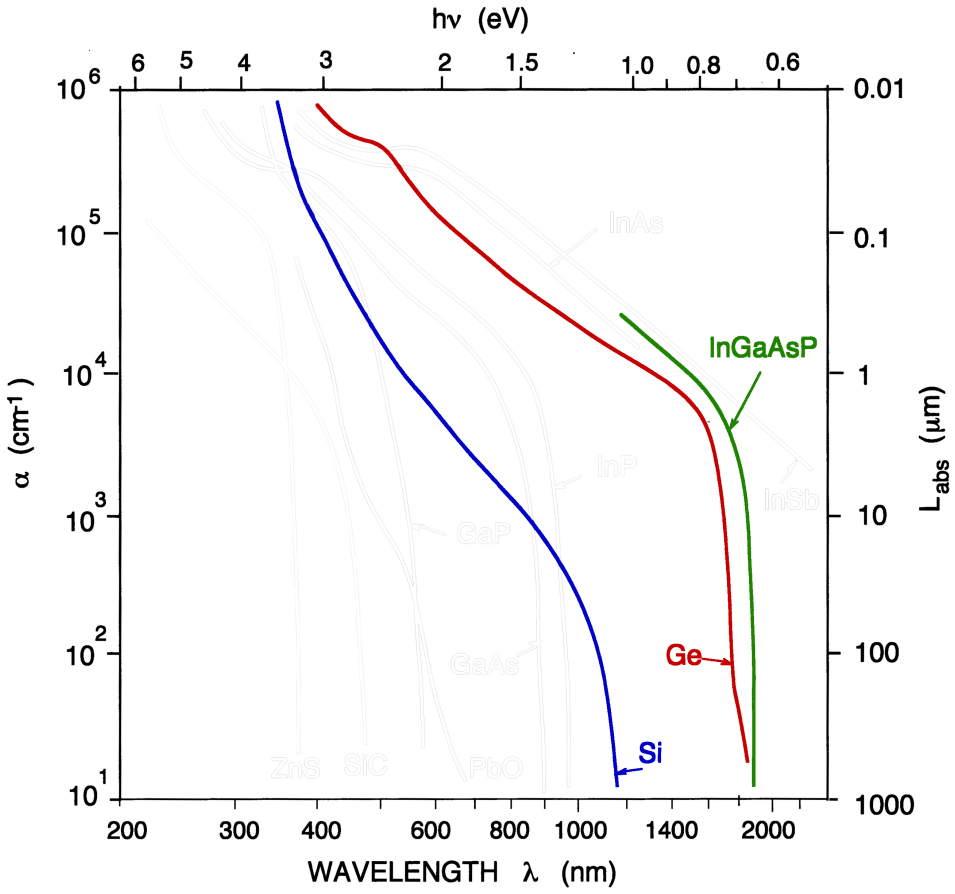

Figure 3.1: Absorption spectrum showing Silicon, Germanium, and InGaAsP, modified from[1]

(HBTs)[72]. This means that not only is it an acceptable material to incorporate into a process for CMOS compatibility, but additionally there is already a wide array of tools and processes which have been developed to grow and deposit SiGe and Ge, interface it with other materials such as silicon, and pattern and etch the material as well. Thus very little additional process development would be needed to incorporate Ge structures.

Because of this, and due to its 4% lattice mismatch with silicon, it has been viewed as a material compatible to be grown on silicon and used as a detector material[75]. There are several different methods to obtain crystalline germanium on silicon substrates. The primary method used is direct epitaxy, where germanium is grown directly on silicon[76]. Because of the lattice mismatch there will be defects, and without proper calibration of the growth conditions, extremely high defect densities, over 10^8 cm^{-2}, can occur. One way to alleviate this is to use a graded SiGe buffer layer, which slowly changes the lattice constant from that of silicon to that of germanium[76], but this can result in a rather thick interlayer. Because the threading dislocations propagate at a nearly 45° angle for $< 100 >$ orientation

germanium and silicon, one method is to selectively growth germanium inside a narrow oxide window[76], where the defects will terminate at the oxide interface on the sides, and the top portion of the germanium will be defect free. This can be extended to create bulk films by patterning the entire substrate with small oxide pits, and growing above the oxide to form a smooth film, referred to as epitaxial overgrowth[77].

Germanium can also be obtained on oxide (GOI) through wafer bonding[78], which is quite similar to the process used to create SOI wafers or even the III-V to silicon bonding used for laser sources in silicon photonics[42]. Another method recently developed is rapid melt growth (RMG) [73, 79], where germanium is patterned on an oxidized substrate and just a small portion of it touches the underlying silicon seed. The germanium is melted and when cooled, the portion touching the seed crystallizes, and since the defects propagate at angles, the defects terminate near the seed, leaving the portion far away from the seed defect free [73]. This is similar to epitaxial overgrowth, and is actually more similar to a micro-scale version of Czochralski growth[73]. Another interesting feature of RMG is that the germanium prior to RMG can be deposited either by CVD or even by evaporation or sputtering[80].

In the methods where germanium is grown directly on silicon, the first 50 nm of material near the interface is generally quite poor and has a large number of defects. Since we are examining how to scale the photodiode or phototransistor to be as small as possible, in a highly scaled device, that first 50 nm will consume a large portion of the device and heavily degrade the performance. From that perspective, the GOI technologies are more attractive, and particularly RMG provides the most flexibility of fabrication.

3.2 Rapid Melt Growth Process

The basic process follows the methodology in Ref [73]. The silicon substrate is oxidized either with thermal oxide or LTO, and a small opening into the oxide is etched with dilute HF to ensure no damage to the underlying crystal. The germanium is deposited via LPCVD followed by a thin LTO layer (10-20 nm) in order to provide stiction for photoresist. The germanium film is patterned into strips 2-20 μm in width and with varying lengths and dry etched using a standard silicon RIE etcher. After etching the germanium is coated in a thick (200-300 nm) layer of LTO and placed into a rapid thermal annealer where the temperature is raised above the melting point of germanium (936 °C) for a second and then rapidly cooled back down. The specific thermal program used is shown in Fig. 3.2. After the wafer has cooled, the melting itself can be verified visually by inspecting areas of the wafer where large sheets of germanium were present, such as in between dies on the wafer. In these areas, the germanium when melted will ball-up, fracturing the top oxide. This will give these portions of the wafer a 'wrinkly' appearance. If the entire wafer has not be melted as determined by this quick visual inspection, the wafer can be re-run through the RMG program without any detriment to the germanium material.

To obtain a quality germanium layer, first a smooth germanium film must be deposited.

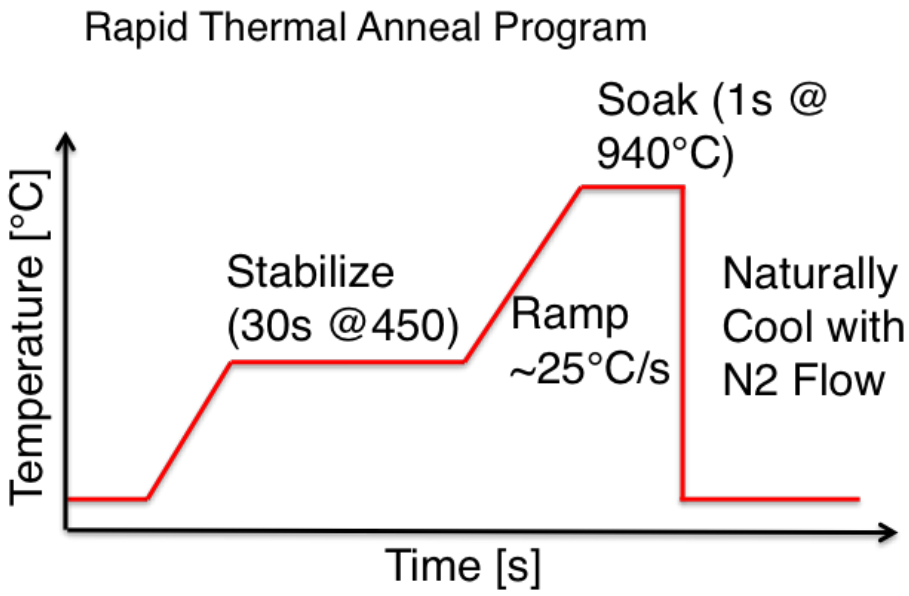

Figure 3.2: Example thermal program to rapidly melt and cool germanium for rapid melt growth.

Because the germanium is melted and recrystallized, the as deposited germanium can be polycrystalline or amorphous, but it should most importantly be pure and smooth. In order to ensure both, the best method with the facilities available was decided to be low-pressure chemical vapor deposition (LDCVD) which deposits a polycrystalline film using MOS grade germane gas and can be done in a batch process on over 25 wafers at a time. The recipe consists of two steps, seeding and deposition steps. Because germanium does not like to wet SiO_2, a silicon seeding layer is needed to ensure stiction. In this case we deposit a roughly 5 nm amorphous silicon film, followed by the germanium. Subsequent characterization does not reveal any degradation of the crystal quality or silicon impurities in the final germanium crystal as a result. The recipe used was taken from a previous user and verified with SEM to view both the film roughness and deposition rate. For the seeding layer we flow 100 sccm of Si_2H_6 at 350 °C for 5 minutes with a process pressure of 300 mTorr. For the deposition we flow 88 sccm of GeH_4 at 350 °C with a process pressure of 600 mTorr. The resulting film grows at roughly equal rates on both silicon and SiO_2 with a growth rate of 6 nm/min. The roughness measured is about 5%.

3.3 Material Quality

It is essential to determine the quality of the resulting crystal before attempting to use it in an electronic device. There are several ways this can be done, including a preliminary visual inspection with either an optical microscope or scanning electron microscope (SEM).

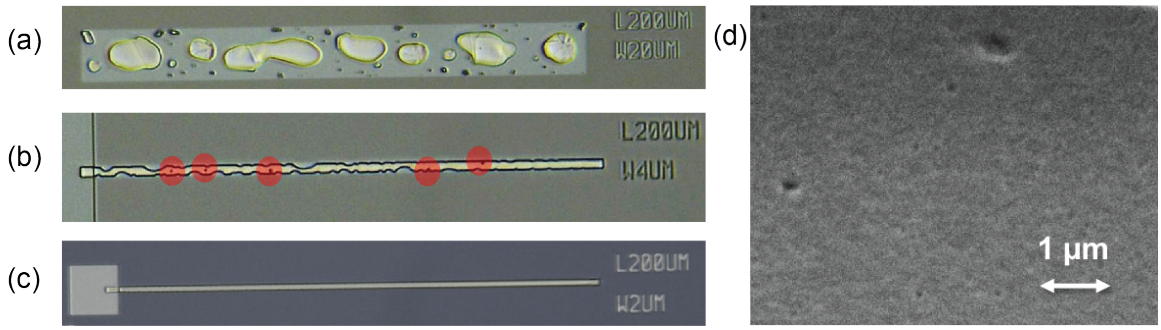

Figure 3.3: (a) Shows a control poly-Ge strip, originally 20 μm with no seed, showing grain boundaries highlighted with the defect etch. (b) A 4 x 200 μm strip of germanium showing 5 highlight dislocations. (c) An example 2 x 200 μm strip of germanium with no threading dislocations, i.e. defect free. (d) An SEM image of a dislocation highlighted by the defect etch, showing a 1 μm diameter pit which would be visible under an optical microscope.

More quantitative methods include cross-sectional transmission electron microscopy (TEM) analysis which can visually pin-point defects and dislocations or high-resolution TEM which can even show crystal structure[73]. Another method to highly defects and dislocations is through the use of chemical defect etching, where the chemical selectively etches defects faster than the bulk crystal, resulting in pits which can be visually identified in a microscope image[81]. Previous works on germanium RMG have used both TEM and defect etching to verify crystal quality resulting in similar quantification of defects[73]. Because of this and the relative simplicity of defect etching, we use a chemical defect etch to quantify the crystal quality.

The chemical defect etch consists of HF:HNO_3:CH_3COOH (1:3:10)[81]. Several etch times were used and the size of the resulting pits were measured under SEM and optical microscopes. Ultimately an etch time of 30 s was used, which resulted in 1 μm diameter pits, as can be seen in Fig. 3.3(d).

Several key features can be seen in Fig. 3.3, the first of which relates to the size of the germanium strip and the resulting shape after RMG. The 20 μm wide strip of germanium is seen to completely deform into several balls of germanium. Even the 4 μm wide strip of germanium can be seen to have some warping on the edges, while the 2 μm strip of germanium nicely retains its shape. This effect has been documented previously for RMG[80], and it is particular to the fact that when liquid, germanium does not like to wet silicon oxide. As such, generally the width of the germanium stripe that can be used for RMG without deformation of its shape is about 3 μm. There is some speculation that using a thicker capping oxide to more firmly contain the germanium while liquid can help to reduce this 'balling up' tendency[80].

From the defect etch, we can first see from the SEM image that a pit is clearly etched, and should be visible in the optical microscope images. Thus in Fig. 3.3(b) we can clearly see 5 threading dislocations which have been highlighted, while in (c) there are no visible dislo-

cations, indicating it is defect free. Looking at many more germanium strips and averaging the number of defects across them, we find that the defect density of the RMG germanium is about 10^6 cm^{-2}, which is as good as high quality epitaxially grown germanium[76]. Finally we can attest that contacting the silicon seed does in fact have an impact on the ultimate crystal structure, since in Fig. 3.3(a) there is no silicon seed, we can clearly see poly-crystalline structure with clear lines marking the grain boundaries. No grain boundaries are seen on any of the RMG germanium strips, implying that the single crystallization extends to at least 200 μm past the seed.

3.4 Wrap-around photodiode design with RMG Germanium

In order to fully characterize the material it is most useful to use it to create a device. The most simple demonstration of RMG germanium is to produce a PIN photodiode, which is shown here. One unique aspect of RMG germanium is its ability to form a crystalline material which is non-planar as seen in the previous SEM images of germanium strips wrapping over the oxide edge from the silicon seed area. This is utilized to achieve a unique device design where the germanium is wrapped around a single mode silicon waveguide, which is shown in Fig. 3.4. The motivation is to create strong coupling between the silicon waveguide like in butt coupling[48] but with the thinner and lower capacitance usually obtained with a thin evanescent coupling design[82]. Additionally, by having the photodiode wrapped around the waveguide, the area just around the waveguide will absorb the most light, meaning the middle intrinsic region will absorb most of the light leaving little absorption in the highly doped contacts on the edges in a lateral PIN scheme. This should improve both the responsively and the speed by reducing the number of slowly diffusing photo generated carriers.

The design has a single mode silicon waveguide (210 x 500 nm) surrounded by a 180 nm thick germanium layer, separated by a thin 20 nm oxide. The oxide serves to as a substrate for the germanium RMG process, it isolates the germanium diode electrically from the silicon, and it protects the silicon waveguide as an etch stop during the germanium etching step of the processing. Thus it should be as thick as possible while not degrading the optical coupling between the waveguide and the germanium.

The p+ and n+ contact regions are on the outer edges of the device and are contacted by a Ti/Al (50nm / 450 nm) layer. By being placed far from the waveguide, the doping and metal contacts interact very weakly with the incident light. This is shown in optical simulation results. The spacing between the doped regions is as wide as 1.1 μm but is also a variable that is made as narrow as 0.5 μm. The distance of 1.1 μm was decided upon such that the doping would only be done on the edge 'fins' but to also be put as close as possible to the vertical edge of the diode as it wraps around the waveguide while leaving 100 nm of lithography tolerance on either side. To make it clearer, if we add the width of the waveguide (500 nm), the 20 nm of oxide on either side of the waveguide, twice the thickness

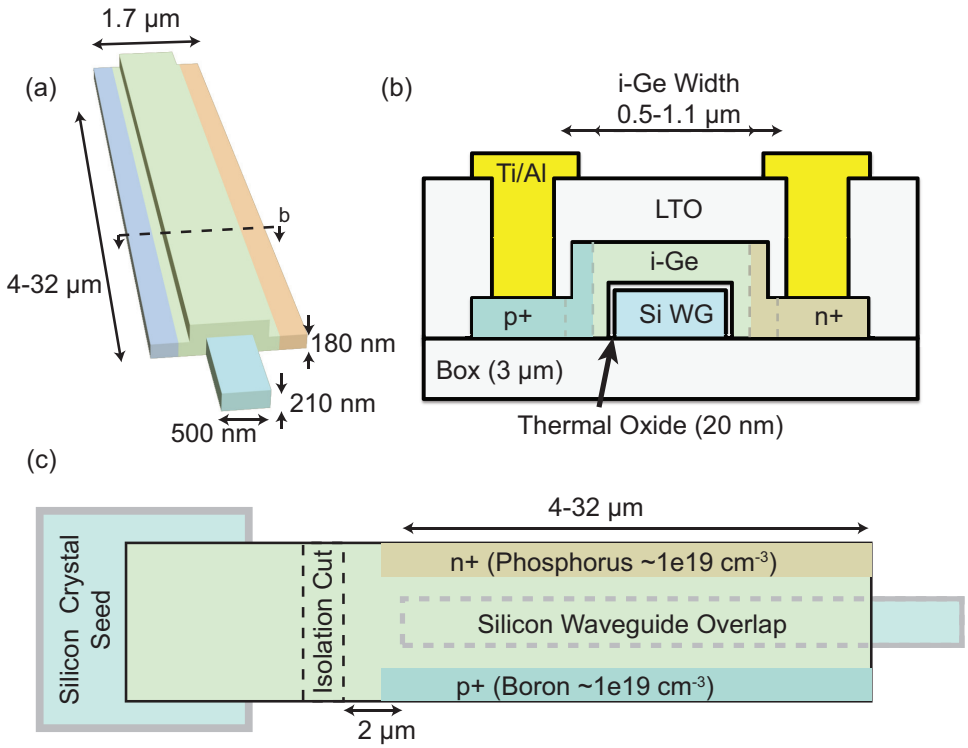

Figure 3.4: (a) 3D Schematic which shows 180 nm thick germanium wrapped around a single mode silicon waveguide with a lateral p-i-n doping scheme. (b) A cross section is shown displaying the p+ and n+ doped contacts on either side of the waveguide with metal contacts. The width of the intrinsic region is a variable which is later characterized. (c) The lateral overview of the device shows the silicon crystal seed with respect to the device and shows the overlap between the germanium and waveguide, as well as the area etched to isolate the photodiode from the crystal seed.

of the germanium itself (180 nm), and add 100 nm of tolerance on either side, we get a total of 1.1 μm. The total width of the diode is 1.7 μm, which comes from the desire to keep the overall width of the germanium stripe to near 2 μm or less to avoid the deformation seen in wider stripes of germanium during the RMG process. This desire for narrower lines needs to be balanced with the need for having enough contact area for low resistance and low RC products.

Additionally, the waveguide overlap is also made variable. It extends 4 to 32μm into the germanium with an additional 1μm of doping going past the end of the waveguide, and then another 1μm going past the end of the doped regions where the device is cut to isolate it from the seeding area. The extension of the device past the end of the waveguide is largely to ensure functionality despite any lithography errors such as over/underexposure of layers and misalignment.

To understand how well the light couples, and to optimize the thickness of the layers,

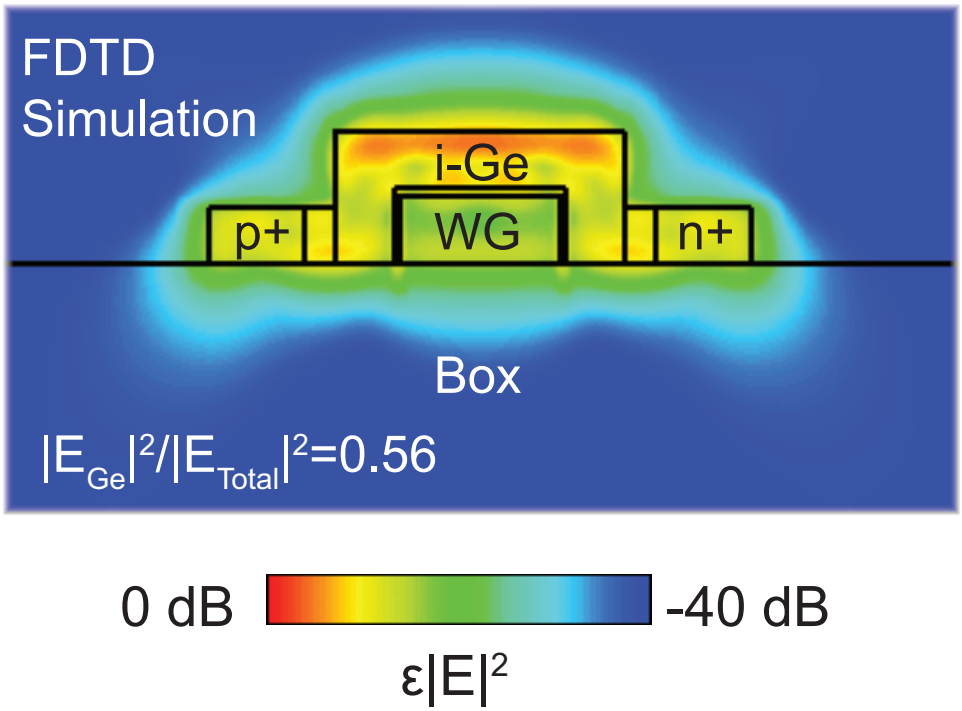

Figure 3.5: Optical energy density from FDTD simulation of the wrap-around photodiode structure with optimized germanium thickness and oxide spacer thickness.

finite difference time domain (FDTD) simulations were performed using Lumerical and CST software packages. The full 3D structure was simulated with different germanium thicknesses, different oxide thicknesses between the waveguide and germanium, and different germanium lengths. The simulated optical energy density of a 2D cross section for the optimized dimensions can be seen in Fig 3.5. What was found was for any thickness of germanium, above 20 nm of oxide thickness, the absorption in the germanium decreased. Thicker germanium resulted in more absorption but with more of the mode present in the side lobes of the device where the contacts and heavy doping would be. In order to be more precise, the absorption in just the i-Ge region was calculated, and it was found that between 150-180 nm was optimal for the germanium thickness. Given 180 nm thick germanium and 20 nm of spacer oxide, the confinement factor of the total energy present in the germanium is calculated, which is $\Gamma_{Ge} = \int |E_{Ge}|^2 / \int |E_{Total}|^2 = 56\%$. Thus the expected absorption for a given length can be approximated using the simple Beer-Lambert absorption formula taking into account the confinement factor and that the absorption length of germanium is about 8 μm. That is for a given length in μm, L, the quantum efficiency for absorption is given by Q.E. $= 1 - e^{-\Gamma_{Ge} L/8}$. For a length of 32 μm we would expect nearly 90% absorption or a responsivity of 1.1 A/W.

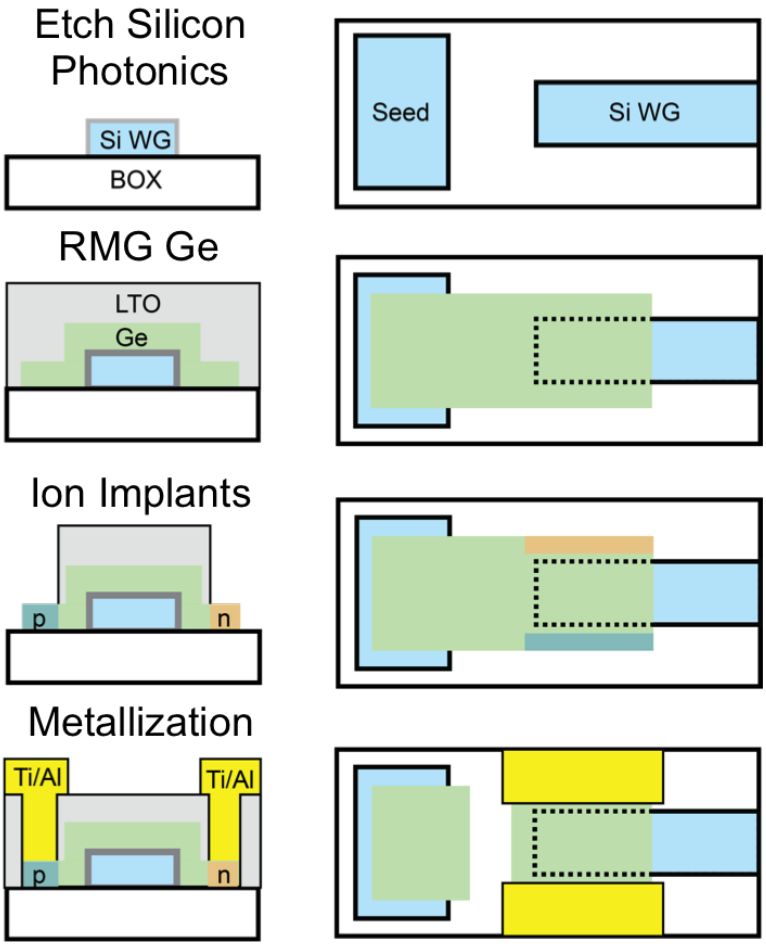

Figure 3.6: Overview of photodiode process flow.

3.5 Wrap-around photodiode fabrication

The basic fabrication process used here remains very similar for all subsequent devices, so a large amount of detail will be given here, while in subsequent sections, largely the key differences will be highlighted. The overview of the process is shown in Fig 3.6.

Starting with a 6" 220 nm SOI wafer with 3 μm box layer, the alignment marks for the ASML stepper are exposed with 20 mJ energy on 400 nm thick DUV (UV-210) photoresist. The resist is UV hardbaked and 120 nm deep trenches are dry etched with an RIE etcher using a standard silicon dry etch recipe. After stripping the resist, the wafers receive a pirhana etch to remove any trace organics, are rinsed, and then dipped in 25:1 HF to remove any native oxide. A 130 nm low-stress nitride is then deposited by LPCVD, which is used as a hardmask during the silicon etching for the photonic structures. The first mask which defines the waveguides and grating couplers is exposed and the nitride is etched using a standard

nitride etch in an RIE etcher. For these critical dimension mask layers, an exposure test matrix is exposed and examined under SEM. The resist mask is stripped, and the wafers are cleaned of any etch polymer residue using an RCA 1 clean followed by an O_2 plasma.

The silicon structures are then dry etched 70 nm deep with a timed standard RIE silicon etch. The etch time is calibrated using a test wafer and the etch depth is measured using a confocal microscope. After the silicon etching, the wafers are again RCA cleaned. The second mask is exposed and hard baked, after which the wafers again go through a silicon RIE etch which etches the remaining 150 nm of silicon down to the box layer. Then the nitride hard mask is removed with 160 °C phosphoric acid. The wafers then go through a final pirhana clean followed by a short 25:1 HF dip and are placed in a thermal oxidation furnace. They are oxidized in ambient oxygen at 1000 °C for 14 minutes, followed by a 5 minute anneal in nitrogen at that temperature. The thickness of 20 nm is verified with an ellipsometer measurement.

The next mask is exposed, which is used to create the openings in the oxide to expose the silicon seeding regions for the RMG. After being developed the photoresist is not hard baked to make it more resistive to wet chemical etchants. A 10:1 buffered HF solution is used to etch the oxide and the rate is calibrated with an unpatterned test wafer. To ensure all of the oxide is gone in the seed area, the wafers are overetched by about 50%. It is important that buffered HF (BHF) is used, because the low pH of unbuffered HF will destroy the photoresist. To avoid oxidizing the exposed silicon the photoresist is removed using an 80 °C PRS-3000 bath instead of the more typical O_2 plasma ashing. The wafers are then quickly placed into the germanium LPCVD furnace where the standard germanium deposition recipe of 5nm of silicon seed followed by 180 nm of germanium are deposited. After removal from the germanium furnace the wafers are visually inspected to verify the film is shiny and silver and not hazy, which is a good indication a smooth germanium film was deposited. The wafers are then immediately placed into a low temperature oxide (LTO) LPCVD furnace where a thin 20 nm LTO layer is deposited. The purpose of the LTO layer is to ensure stiction of later photoresist layers on the germanium.

The next photomask is exposed which defines the germanium stripes. The photoresist is UV hardbaked after development, and first a short oxide RIE etch is performed to etch through the 20 nm LTO layer. Then a standard silicon RIE etch recipe is used to etch through the germanium, which when calibrated was found to etch germanium about 3 times faster than silicon. In order to clear most of the germanium, a slight amount of overetching is performed, but etching through the thermal oxide protecting the underlying silicon waveguide should be avoided. Generally about 5-10 nm of thermal oxide is consumed with 5-10 s of overetching.

Because of the topography present in the design, there will be some germanium 'stringers' left near the sidewalls of waveguides or other silicon structures. This can cause undesired loss along the waveguides by providing structures to scatter the light as well as the fact that the germanium will absorb the light. To remove this we have another mask which covers the germanium structures with a 500 nm extension beyond the germanium. This photomask is again not hardbaked after development to make it more resistant to wet chemical etchants.

Additionally, the photoresist covering the alignment marks is manually removed using a cotton swab and acetone. The wafers are then dipped in 10:1 BHF to remove the top 20 nm LTO layer covering the germanium. A test wafer verifies the etch rate. Then the wafers are placed in a 50 °C bath of 30% H_2O_2, which removes the germanium. Since the germanium has a silver color and the underlying silicon appears slightly blue, the etch completion can be verified by a visual color change. To clean up any remaining residue, the wafers are again placed in the 10:1 BHF for a few seconds. The goal of this step is to remove the germanium on the backside of the wafer, the germanium covering the alignment marks, and the remaining germanium along the waveguide sidewalls. During the RMG process the germanium in bulk layers will ball-up, which when many germanium beads cover the alignment marks, subsequent alignment steps can be difficult or fail altogether since the alignment is automated. Having germanium on the backside of the wafer during the melt could melt and drip from the bottom of the wafer causing damage to the rapid thermal annealer, in addition to adding roughness to the wafer backside which could also impair subsequent lithography steps.

Now the wafers are again placed in the LTO LPCVD furnace and 400 nm of undoped LTO is deposited to completely encapsulate the germanium for the RMG process. After the LTO is deposited the wafers are immediately placed in the rapid thermal annealer which runs the RMG program, 1050 °C for 1 s. The wafers are visually inspected to make sure the entire wafer was melted by examining the lines of germanium running between dies and looking to see that they look 'wrinkly'. If the entire wafer has not fully melted, it is ok to run the melt program again.

The next lithography mask exposes the contact hole openings in the oxide, which after a UV hardbake, the openings are etched with an oxide RIE process. An unpatterned test wafer is used to calibrate the etch rate and some overetching is performed. Because the selectivity of the oxide etch against germanium is not perfect, some of the germanium will be etched. After the contact holes are etched, the two implantation masks are exposed separately, though the order is not important. The n-implant used is a 15 keV phosphorus implant at 7 ° of tilt with a 1E15 cm^{-2} dose. The p-implant is 25 keV boron at 7 ° tilt with a 1E15 cm^{-2} dose. The implants are chosen to provide a relatively shallow doped region (150 nm) with moderately high carrier concentration (1E19 cm^{-3}) for good contact resistance. Both implants are performed by an outside vendor since our lab does not have an ion implanter. After the implantation, the wafers are activated in a rapid thermal annealer at 450 °C for 30 s. A lower temperature and shorter time are designed to avoid too much phosphorus diffusion in the germanium[83].

A quick rinse in water should remove any native oxide formed on the germanium before the wafers are placed into the load-lock of the metal sputtering chamber. To ensure good contact, a quick sputter etch is performed in the chamber to remove the first few nm of material before depositing the metal. A 50 nm Ti layer is first sputtered, followed by 450 nm of $AlSi_{2\%}$. The final lithography mask is exposed which defines the metal contacts. After a UV hardbake, the wafers are etched in an aluminum RIE etcher and several seconds of overetching are performed to ensure all of the metal, including that along the sidewalls of

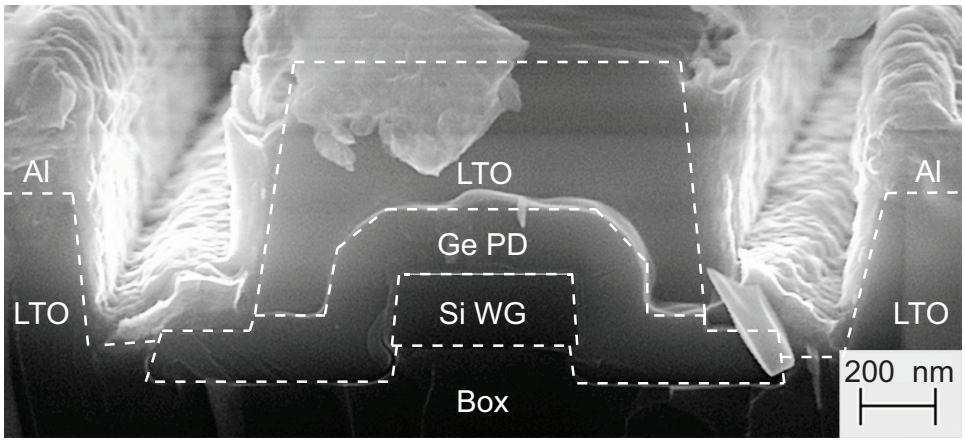

Figure 3.7: Cleaved cross-section of final fabricated wrap-around photodiode imaged with SEM.

the waveguides is removed. A completed device was cleaved and the cross-section imaged by SEM, which is shown in Fig. 3.7.

3.6 Wrap-around Photodiode Characterization

3.6.1 DC Characterization

In characterizing the photodiodes, both the optical and electrical properties need to be characterized. Since the photodiodes are on waveguides, the insertion loss of the waveguides needed to be characterized. All of the photodiodes were designed for measurement by a fiber array. All devices were arranged in a row with the grating couplers separated by the pitch between fibers in the array. Both the fiber array and the array of waveguides and devices can be seen in Fig. 3.8 The outermost gratings were connected to each other by a waveguide so that the optical insertion loss could be measured during the alignment of the fiber to the grating with the device. This ensures that for each device measured, the real grating insertion loss is known. A 1550 nm laser source was fed into a polarization controller, which allowed the polarization to be matched to that of the grating coupler input. The polarization controller output was then fed into the fiber array which consists of polarization maintaining (PM) fibers. The output fiber on the fiber array was then fed into an optical power meter.

The insertion loss was measured by first determining the optical power of the source by feeding the output of the polarization controller to the optical power meter, and this power was set as the 0dB point. The polarization controller output was then put into the input fiber in the array, and the array was then carefully aligned with a 6-axis stage until the total power throughput was maximized. The polarization controller was adjusted as well to maximize the throughput. The total insertion loss of the grating coupler to grating coupler structure was then divided by two and that value is taken as the grating coupler insertion

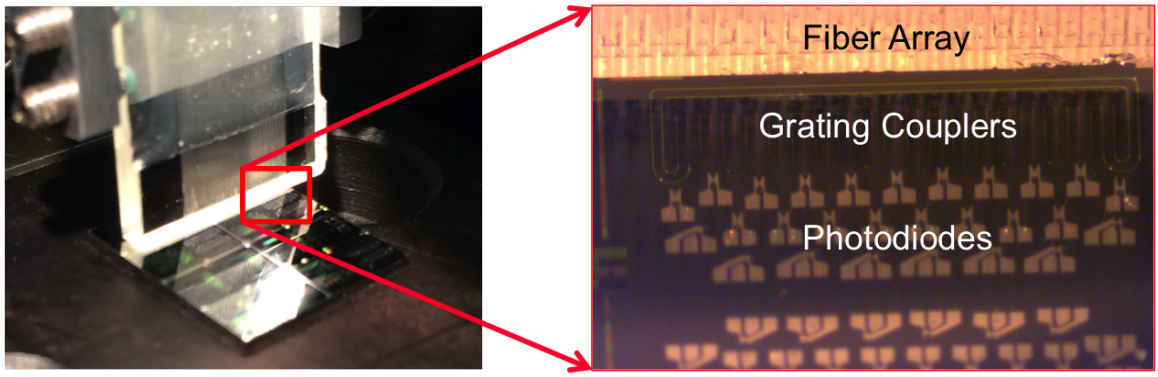

Figure 3.8: Image showing the fiber array coupled to the fabricated chip, with the zoomed-in photo showing the array of fibers aligning with the array of waveguides, which are connected to the photodiode devices. The outer waveguides can be seen to connect to each other, providing information on insertion loss, and allowing the fiber array to be properly aligned to all of the waveguides.

loss. Separate structures with varying lengths of waveguide were also measured to determine the waveguide loss as a function of length. Because all of the waveguides connecting the photodiodes were less than 3 mm in length, the waveguide loss is considered negligible. At 1550 nm, the grating coupler loss was determined to be about -7.1 dB on average across multiple dies. This insertion loss is taken into account in all future responsivity calculations such that the responsivity is normalized to the amount of power incident on the photodiode itself, not the power being fed into the grating coupler.

The electrical response of the photodiode is then measured by probing the diode with DC tungsten electrical probes, and sweeping the voltage across a reasonable range with an HP 4145B parameter analyzer, and measuring the resulting current. This was done first with the microscope lights and laser off to measure the reverse leakage or dark current of the photodiode. Then the laser was turned on and different amounts of optical power were applied to get the photocurrent curves for the diode. An example diode is shown in Fig. 3.9(a).

This particular device is 16 μm in length with a 1.1 μm wide depletion region. The dark current, I_{dark} at -1 V is only 6 nA and even at -5 V of bias is less than 300 nA, which both imply very good material quality. In fact if we calculate the dark current density at -1 V using the cross-sectional area of the junction, it is less than 200 mA/cm^2, which is comparable to germanium diodes made with selective epitaxy[48, 49]. When the photocurrent, I_{photo} is calculated with respect to the power incident, P_{inc}, on the photodiode, we can calculate the responsivity, $R = (I_{photo} - I_{dark})/P_{inc}$, which is shown in Fig. 3.9(b). We can see the responsivity is consistent over a wide range of optical power levels, meaning it is linear over at least a 17 dB power range. Additionally we see the responsivity is 0.95 A/W at -2 V, and

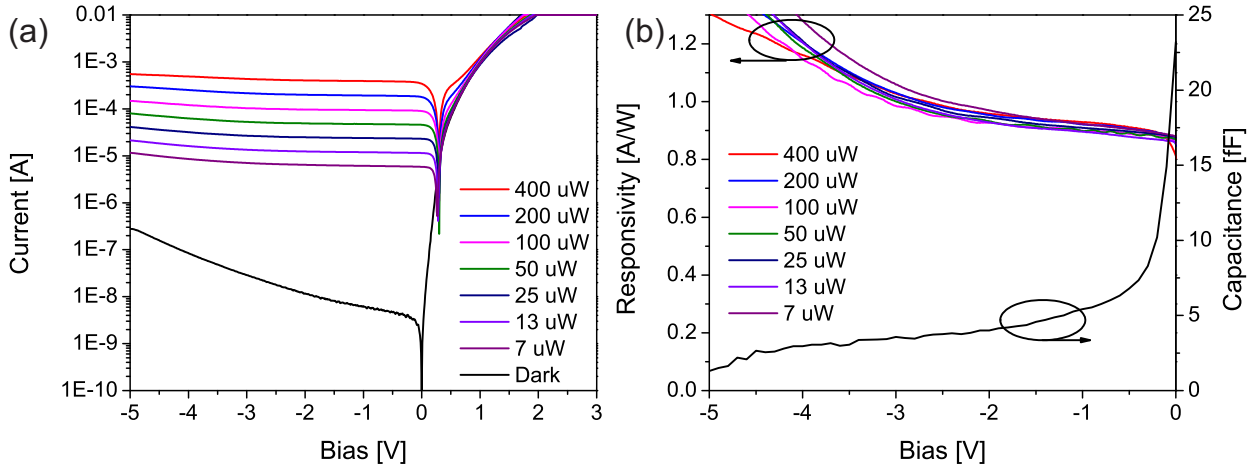

Figure 3.9: (a) The current-voltage relationship for a 16 μm long diode with 1.1 μm wide intrinsic region. Curves are shown both for dark conditions and illuminated with 1550 nm light. (b) The calculated responsivity vs bias for the same device, and the measured capacitance vs bias.

increases slowly from 0 V to -5 V of bias. The responsivity increases dramatically from -3 V to -5 V and it is likely that here the device is fully depleting, collecting all of the carriers that are generated in the intrinsic region. It is also possible that at these voltages due to the small size of the device that the increase in responsivity is also the onset of mild avalanche gain, since the avalanche breakdown field for germanium is 10^5 V/cm.

Another key aspect of the photodiode is its capacitance. This is important for two reasons, one of which is that we chose this particular design to minimize the overall capacitance of the diode. Additionally we can infer the amount of depletion in the intrinsic region by measuring the capacitance as a function of bias voltage. As we apply a large reverse bias the diode will further deplete, reducing its effective capacitance until it is fully depleted, at which point the capacitance will remain mostly constant with respect to voltage. We measure the capacitance using an LCR meter with a test bias of 100 mV at 1 MHz. The results for the same diode are shown in Fig. 3.9(b). We can see that the capacitance drops dramatically until about a -2 to -3 V bias at which point it remains fairly stable at about 4 fF, implying the fully depleted condition is around -3 to -5 V with a capacitance of 4 fF.

This agrees well with the responsivity as a function of bias and to further confirm, electrical simulations using TCAD Sentaurus were performed to examine depletion width as a function of bias assuming varying amounts of background p-type doping in the intrinsic region. It is found that by assuming a background concentration of 10^{16} cm^{-3}, the diode will not fully deplete until a 4 to 5 V reverse bias is applied. Thus the responsivity and C-V measurements also agree with the previous assertion of material quality being high enough to have a low background concentration of 10^{16} cm^{-3}.

It is also important to measure the effect of responsivity as a function of the length of the

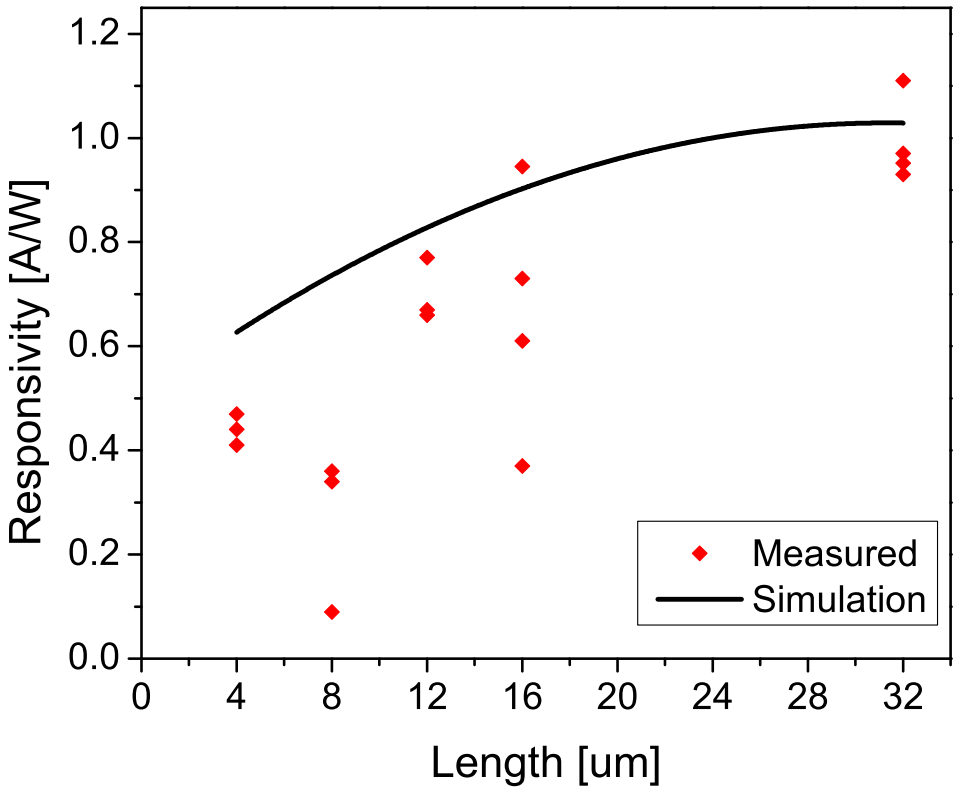

Figure 3.10: Measured responsivity at -2 V and 1550 nm light for multiple diodes at various lengths overlaid with the simulated responsivity as a function of length.

diode. This gives a more complete story of the absorption of the light by the germanium. Fig. 3.10 shows the result of measuring the responsivity at -2 V for multiple photodiode of different lengths. Superimposed is the line representing the simulated responsivity assuming pure germanium. For the most part, the measured responsivity closely matches the simulated value, which implies the RMG germanium has a nearly identical absorption coefficient to that of bulk germanium. For longer and shorter diodes, the responsivity is very similar across multiple devices, indicating good device uniformity. For intermediate lengths, e.g. 8-16 μm, the likely culprit of the increased variation in the dimensions or edge roughness of the germanium. Because the simulations at those lengths show some resonant enhancement, any variation in length or roughness would inhibit that resonance resulting in a dramatic reduction in responsivity. It should also be mentioned that the dark current at -1V for all of these diodes was measured as well, and while a few devices show dark currents in the nA range, the more typical average was a few μA of dark current. This could be due to inconsistent crystallization with the RMG process, and likely a hydrogen anneal after the RMG step or additional surface passivation would reduce the dark current overall[84].

3.6.2 RF Characterization

To measure the high frequency response of the photodiodes a network analyzer (Agilent PNA) was used to record the S_{21} parameter. To do this, port 1 of the PNA drove an EO Space 40 GHz intensity modulator, which was separately biased with a DC voltage to its 3dB crossing point. A tunable fiber laser first went through a polarization controller to align its polarization with the modulator was fed into the input of the modulator. The modulator output then went into an erbium fiber amplifier (EDFA) to boost the optical power level as needed. The output of the EDFA went into another polarization controller to now align the polarization with that of the grating coupler on the chip. The fiber was aligned to the grating coupler using a piezo actuated stage which moved a holder designed to hold the single-mode, cleaved fiber at 8° with respect to normal from the surface. The photodiode is probed with a GGB 50 GHz G-S probe, and the RF signal is input to port 2 of the PNA. The DC bias for the diode is applied through the PNA which has an internal bias-T. Because of the limitations of the modulator, traces up to only 40 GHz were taken. For these measurements, an incident optical power of about 100 μW was applied to be incident on the photodiode.

An important aspect of the measurement is the calibration to remove any frequency response in the system other than that of the diode under test. To do this, a 70 GHz Finisar photodiode was directly connected to the output of the EDFA and its RF port was fed into port 2 of the PNA. A through calibration was performed to remove the frequency response of the system. The Finisar photodiode has a 1 dB drop at 40 GHz that is roughly linear over frequency. To counter the fact that this was calibrated out, when taking traces of the real devices under test, we need to add back the 1 dB of loss over 40 GHz, which is done in post-processing of the S_{21} curves.

For the photodiode mentioned in the previous section, its 3dB bandwidth at -5 V was found to be 9 GHz. The 3dB was extracted form the S_{21} curve by fitting with a polynomial and taking the 3dB point from the fitted polynomial. What might be expected from this diode with a 1.1 μm i-Ge width is that the bandwidth might be higher for lower voltages since the field at -1 V across that distance is 10^4 V/cm. However there is likely slow carrier diffusion from undepleted parts of the diode since in the previous section, simulation analysis showed that full depletion was likely not occurring until -5 V of reverse bias. However at -5 V simulations do show that the electric field is uniformly strong in all parts of the depletion region, above the threshold of 10^4 V/cm needed to obtain velocity saturation of carriers in germanium. The expected speed then at -5 V is 13.5 GHz, which takes into account a total travel distance of 2 μm including the horizontal and vertical components in the travel of a carrier from one side to the other of the diode, and a saturated carrier velocity of 6x10^6 cm/s. However the calculation of 13.5 GHz comes from the formula assuming uniform carrier generations[61], $f_{3dB} = 0.45\frac{v_{sat}}{W_{dep}}$. In this case we know from the optical mode present in the FDTD simulation shows absorption is not uniform, and thus the carrier collection speed would be affected by this as well.

To increase the bandwidth of the photodiode, one must simply decrease the transit dis-

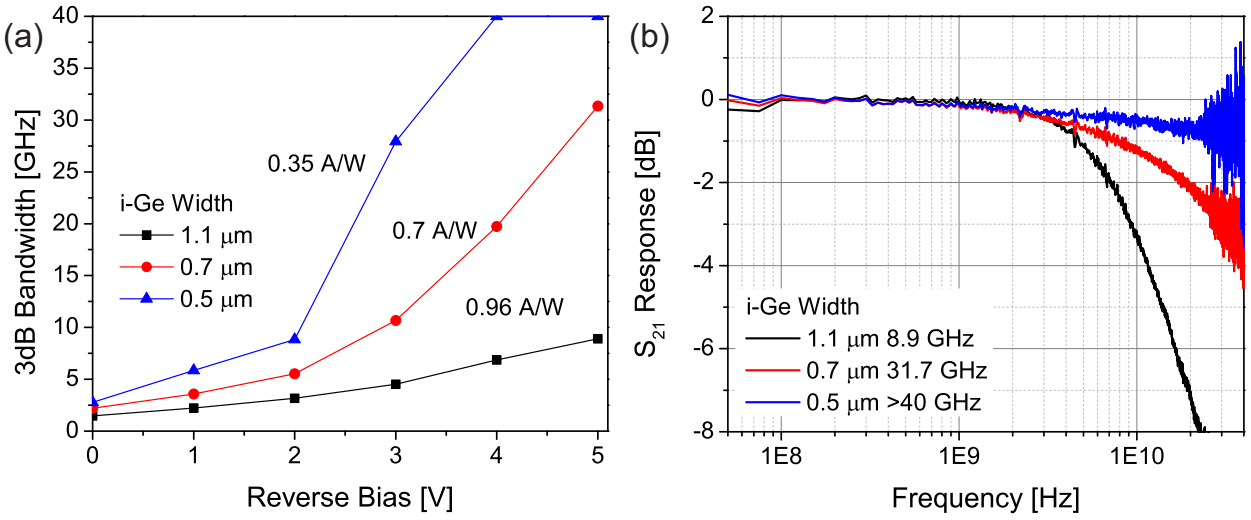

Figure 3.11: (a) Measured 3dB bandwidth for three diodes with different i-Ge widths as a function of bias. (b) S_{21} spectra of those three photodiodes at -5 V bias.

tance since these photodiodes are transit limited. If one extracts the series resistance from the forward bias portion of the IV curve, the obtained resistance is about 25 ohm. Combined with the measured capacitance of 4 fF, the RC limited bandwidth is 1.6 THz. Thus to decrease the transit time, the i-Ge width should be reduced, which was done on several diodes, shrinking it from 1.1 μm to 0.7 and 0.5 μm. While the responsivity is reduced slightly when shrinking the i-Ge width, the bandwidth increases dramatically. In fact by reducing the i-Ge width of 0.5 μm the responsivity is only reduced to 0.35 A/W, and this is because the optical mode primarily resides in the center of the i-Ge region above the waveguide. Thus very little of it overlaps with the heavily doped p and n contacts even when the i-Ge width is narrow. The benefit is that the reduced transit distance greatly increases the bandwidth to beyond 40 GHz. In fact it can be seen that with a -5 V bias, at 40 GHz there is only a 1 dB drop in the S_{21}, suggesting the true bandwidth could be much higher. The calculated bandwidth for the 0.5 μm width i-Ge photodiode assuming velocity saturation is 54 GHz, which seems very reasonable given the measure S_{21} spectra.

3.7 Comparison with similar photodiodes

We can examine the quality of both the material and the resulting device by comparing its key metrics against existing reports in the literature. This is summarized in Table 3.1. It can be seen that in terms of dark current, responsivity, capacitance, and ultimate bandwidth, the photodiode produced with RMG is comparable or better than currently reported photodiode devices. This encourages the use of RMG germanium for the more advanced processes in the following chapters as a material for phototransistors.

Table 3.1: Summary of recently reported germanium waveguide photodiodes.

Group	Growth	Responsivity [A/W]	Dark Current	f_{3dB} [GHz]	L x W x t [μm]	Capacitance [fF]
LETI[48]	Epi	1	18 nA	42	15 x 3 x 0.43	19
LETI[49]	Epi	0.8	4 μA	120	10 x 10 x 0.4	71
Kotura[50]	Epi	1.1	240 nA	32	10 x 0.8 x 2.4	4.8
Kotura[51]	Epi	0.8	500 nA	8.3	200 x 3.5 x 0.9	122
IBM[52]	RMG	0.14	50 μA	40	30 x 0.5 x 0.1	10
Sandia[53]	Epi	0.8	3 nA	45	4 x 1.3 x 0.6	1.2
Jilin Univ.[54]	Epi	0.37	310 nA	20	5 x 3 x 0.1	0.14
Oracle[55]	Epi	0.8	400 nA	15	10 x 2 x 0.3	0.6
A*STAR[56]	Epi	0.65	60 nA	18	20 x 2.4 x 0.2	0.8
Tsing Hua U.[57]	RMG	0.7	7 nA	17	15 x 2 x 0.3	7
This Work	RMG	0.95	6 nA	9 (40)	18 x 1.7 x 0.2	4

Chapter 4

Germanium Gated PhotoMOSFET

The reduction in receiver energy/bit comes from reducing the photodiode capacitance, and reducing the input capacitance of the transistor. With fF photodiodes, and sub-fF transistors, one begins to consider even the capacitance of the metal line connecting the photodiode to transistor with a capacitance of 0.2 fF/μm, almost independent of geometry [7]. This implies that 5 μm of wire connecting a 1fF photodiode to the receiver circuitry effectively doubles the diode capacitance, while it takes only 1 μm of connecting wire to double the capacitance of a 200 aF photodiode. To solve this issue, the wire between photodiode and first transistor should be eliminated by integrating gain with the photodiode.

Several groups have recently demonstrated devices such as a normal incidence germanium photo bipolar junction transistor (BJT) [68], a waveguide integrated germanium junction field-effect transistor (JFET) [67], and germanium avalanche photodiode [4, 62], and a normal incidence germanium gate MOSFET [10]. One issue with the previous photoMOSFET was that a polycrystalline germanium gate was deposited, which greatly inhibited performance.

In this chapter we demonstrate a waveguide integrated germanium gate MOSFET, where RMG is used to create a monocrystalline gate. The NMOS device is fabricated using a self-aligned process on a 220nm-thick SOI substrate that is compatible with standard silicon photonics integrated circuits. The source/drain activation anneal step melts and recrystallizes the germanium. A high responsivity of 18 A/W at 583 nW of 1550nm light with only 8 μm of germanium along the waveguide indicates strong internal transistor gain. Due to a relatively long (1 μm) channel, the device has a 3dB cutoff of 2.5 GHz, obtained with higher optical power. Standard scaling methodology is expected to provide for improvements in both the speed and responsivity of the device.

4.1 Basic Theory of Operation

The device is a photodiode with the cathode attached to the gate contact, and the anode attached to the gate oxide of the transistor. When not illuminated, the diode is reverse biased, and no current can flow through it, so there is a negligible voltage drop across the

diode. In this case, the device functions much like a traditional long-channel MOSFET with a partially depleted gate, since the doping in the germanium gate is not very high.

When illuminated the germanium gate acts like an open circuit solar cell [61] and the photocurrent is balanced by recombination currents, resulting in a net photovoltage across the germanium. This voltage adds to the voltage applied by the gate contact, and the net voltage on the gate is higher. If we assume the germanium has a quantum efficiency, η, then the photovoltage, V_{photo}, can be described by Eq. 4.1 as a function of incident optical power. Here P_{inc} is the incident optical power on the device at a frequency, ν, and I_s is the diode leakage current.

$$V_{Photo} = \frac{nKT}{q} ln\left(\frac{P_{inc}}{I_s}\frac{\eta q}{h\nu}\right) \tag{4.1}$$

Much like the subthreshold slope of a MOSFET, in this device, the additional gate voltage induced by light absorption follows a slope of 60 mV/decade of optical power, assuming Shockley-Read-Hall recombination dominates. The total voltage induced however, is also influenced by the quantum efficiency of the absorption and the internal diode leakage current. This condition will have an impact on ultimate device performance.

From this we can then calculate the photocurrent in the drain current, I_d, when the device is in saturation using the simple square law current equation. This gives both the bias current from the gate and drain voltages and the photocurrent in Eq. 4.2. Here it is assumed the transistor has gate length, L, gate width, W, with electron mobility, μ_n, gate oxide capacitance, C_{ox}, a threshold voltage, V_T, and an applied gate bias of V_g.

$$I_d = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_g + V_{photo} - V_T\right)^2 \tag{4.2}$$

It can be seen the photocurrent has a ln^2 dependence on incident optical power as well as a dependence on gate voltage. Because the device is fundamentally an NMOSFET with a light modulated gate, one can choose which region to operate in depending on the requirements of the device. Much like an amplifier, this implies that in order to get a larger photoresponse from the device, a larger bias current is needed.

4.2 Design and Fabrication Process

4.2.1 Design

Shown in Fig. 4.1(a) is the schematic cross-section of the device. On the surface, the device is a traditional self-aligned NMOS with a silicon channel and a germanium gate. The silicon body layer is 220 nm thick, on 3 μm of buried oxide, and is also used as the silicon photonics layer. Coming out of the plane of the figure is the single mode silicon waveguide (500 nm wide) which tapers out to a grating coupler for optical input. The single mode waveguide abuts the transistor body as seen in the microscope image in Fig. 4.1(b). The light in the transistor body is then evanescently coupled into the germanium gate due to the higher

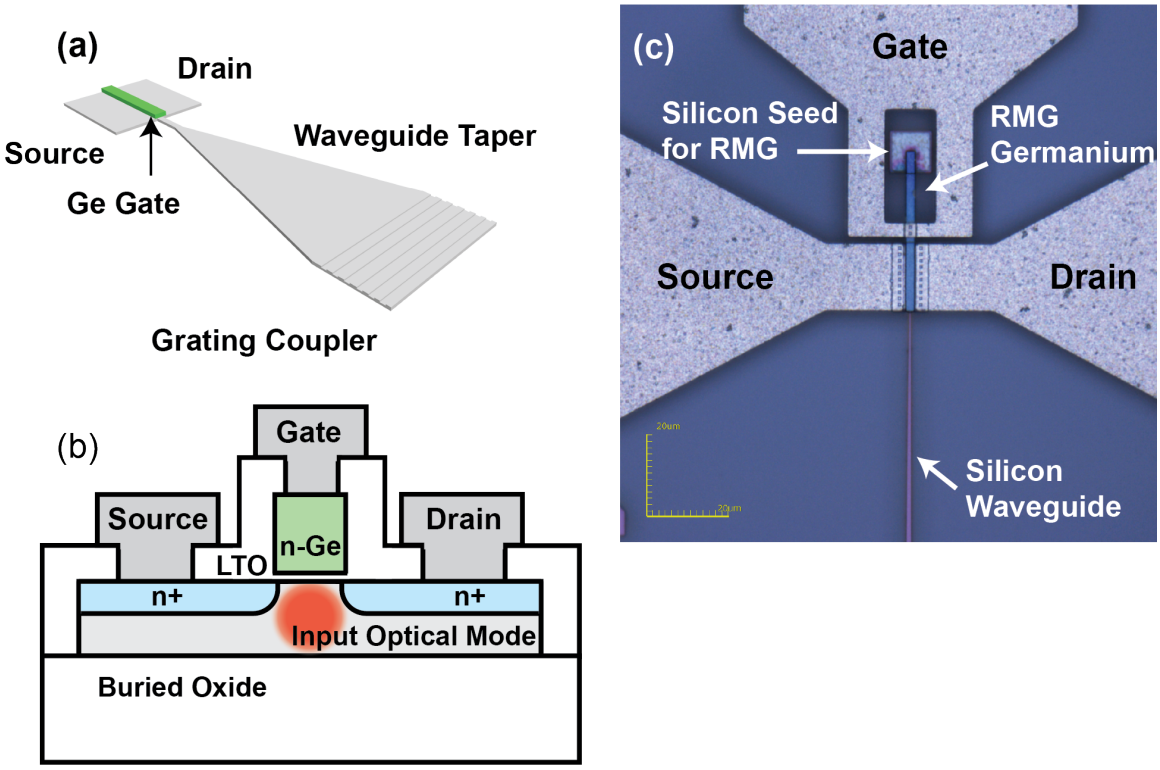

Figure 4.1: (a) 3D Schematic drawing of the photoMOSFET with integrated silicon photonics components (b) Schematic cross-section of the photoMOSFET (c) Optical micrograph of a completed photoMOSFET.

refractive index of germanium [74]. Because the absorption length of germanium at 1550 nm is 5 μm, the gate widths were designed from 2 - 32 μm, with channel lengths between 0.25 - 2 μm. The confinement factor in the 1x8 μm germanium gate is 42.6% according to finite difference time domain (FDTD) mode calculations, which makes effective absorption coefficient 852 cm^{-1}. For 8 μm of interaction length, in a single pass there should be 49% absorption in the germanium gate. An FDTD simulation estimates 55 % absorption in the gate including back reflections. Fig. 4.2 shows the energy distribution both in the germanium gate and the underlying silicon from FDTD.

Because the bandgap of silicon is 1.1 eV and is indirect, the primary absorption in silicon at 1550 nm is due to free carrier absorption [85], which is minimized with a low p-type body doping of 10^{17} cm^{-3}. The silicon waveguide and all other silicon photonics structures have a much lower doping level of 10^{14} cm^{-3}, which will give low propagation loss. Additionally the mode quickly couples into the germanium gate, avoiding optical interaction with the much more heavily doped source/drain region in the silicon. Also seen in Fig. 4.1(c) is an electrically and optically isolated island of silicon near the transistor body which is used as the seed to crystallize the germanium during the RMG process [73]. Because the silicon

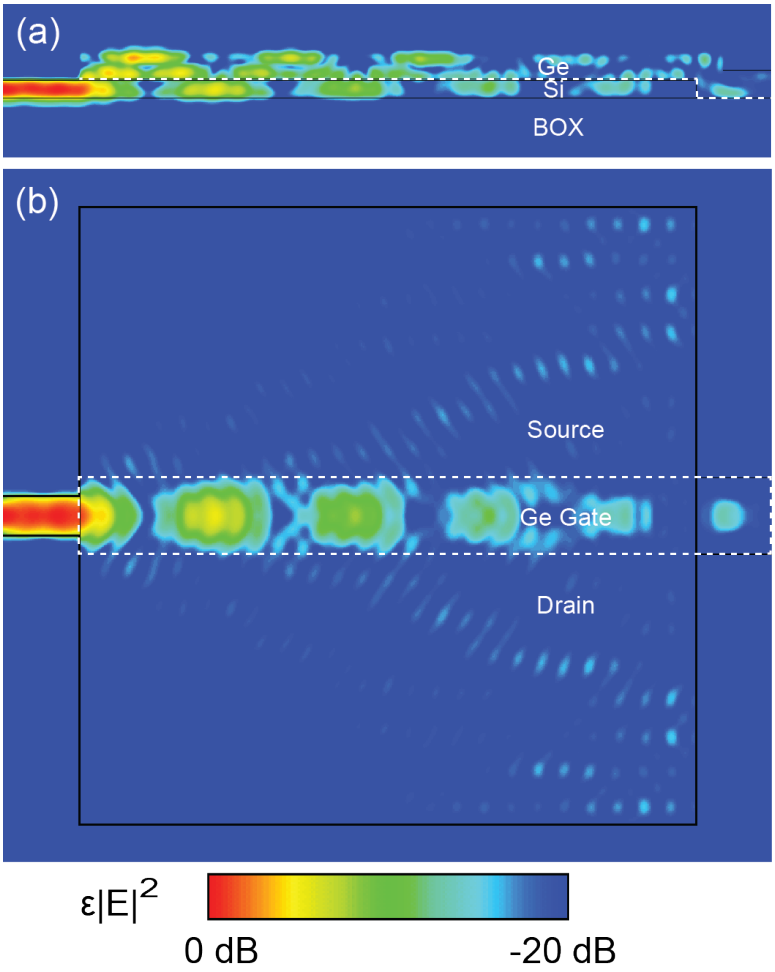

Figure 4.2: (a) Slice through the middle of the gate, showing the energy distribution from FDTD between the germanium gate and silicon body along the interaction length. (b) Slice through the body of the transistor, showing energy confinement under the germanium gate in the silicon.

body and crystal seed are separated laterally by the buried oxide layer, they are electrically isolated from each other.

The body doping was chosen to be uniform for simplicity with a single boron implant giving a channel doping of 10^{17} cm^{-3}, which would give a maximum depletion width of 105 nm [61]. The source and drain doping were chosen to give 10^{20} cm^{-3} n-doping for ohmic contact, while simultaneously n-doping the germanium gate to 10^{18} cm^{-3}. A tailored screening oxide is deposited over the germanium to control the dose implanted into the germanium, allowing a lower doping concentration than the source and drain regions. A moderate doping concentration in germanium is desired to strike a balance between lowering the amount of depletion in the gate and having a lower recombination rate in the germanium for an optimal photoresponse.

4.2.2 Fabrication

Starting with a 220 nm SOI substrate with 3 μm buried oxide layer, a blanket boron implant was performed with a 10^{12} cm^{-2} dose at 30 keV with 7° tilt. The grating couplers were then formed with a 55 nm etch into the silicon (70 nm designed), followed by a full 220 nm silicon etch to form the waveguides. The full etch also isolates the crystal seed from the waveguides and transistor body. The wafers were then cleaned in piranha solution (1: 50 H_2O_2:H_2SO_4) followed by a 10:1 buffered HF (BHF) dip, to remove any organics and native oxide present. The wafers were then thermally oxidized in dry O2 at 950 °C which grew a measured 17 nm oxide. The oxide was annealed in N_2 at 950 °C for 30 minutes.

A 10 μm seed hole was opened in the oxide with a dilute (50:1) HF solution after being defined lithographically. The resist was then immediately stripped with PRS-3000, cleaned in piranha solution followed by a 1 s 10:1 BHF dip, and placed under vacuum in a germanium LPCVD furnace. The gate oxide was measured to be 16 nm after the cleaning step before germanium deposition. To promote adhesion to the oxide, a 5 nm amorphous silicon layer was first deposited, followed by a 350 nm polycrystalline germanium layer. Finally a 20 nm LTO layer was deposited on top of the germanium to promote adhesion for subsequent photoresist steps. The thickness of this LTO also determines the total implant dose in the germanium.

The germanium gate was then patterned and dry etched. A self-aligned phosphorus implant, with a dose of 10^{15} cm^{-2} at 20 keV with 7° tilt then formed the source, drain, and gate doping. 300 nm of LTO was deposited on the wafer, followed by a backside etch in HF and H2O2 at 50 °C to remove germanium from the backside of the wafer. Source/drain dopant activation was achieved with a rapid thermal anneal at 1000 °C for 1 s. This step also melted and recrystallized the germanium as in the RMG technique. Because the germanium is liquefied in this step, it is assumed the phosphorus dopants distribute uniformly throughout the gate.

Finally contact vias were formed in the oxide to the germanium and silicon by lithographic patterning and dry etching. The wafers then received a quick (10:1) HF dip and were immediately placed into the load-lock of an RF sputterer. 50 nm of Ti followed by 400 nm

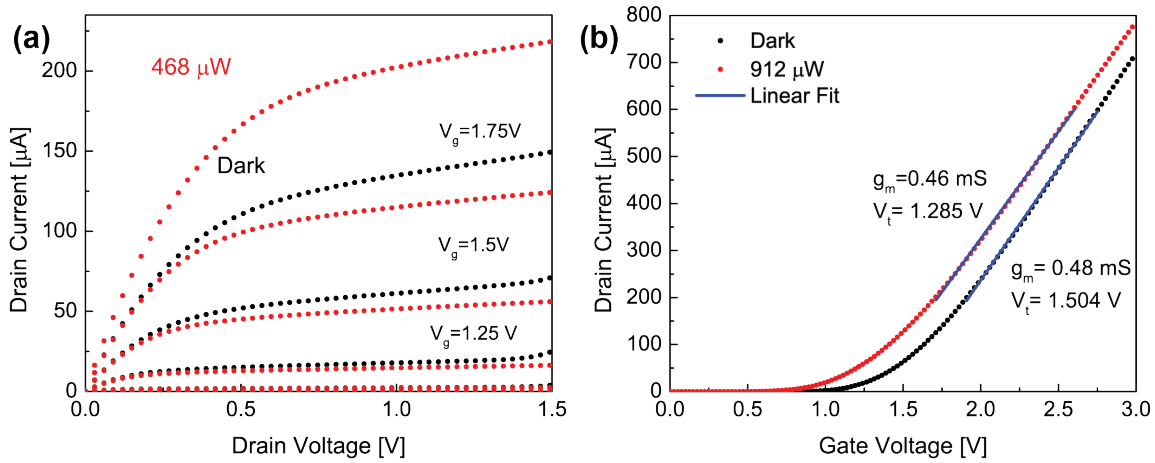

Figure 4.3: (a) Drain current as a function of drain and gate voltage both in the dark and with 468 μW of illumination. (b) Drain current as a function of gate voltage at a drain voltage of 1 V. The transconductance and threshold voltage are extracted both in the dark and with 912 μW of illumination.

of $Al_{.98}Si_{.02}$ were sputtered onto the wafer. The metal layers were patterned and dry etched to form contact pads, and the wafer was then annealed in forming gas (H_2/N_2) at 450 °C for 30 minutes to form ohmic contacts.

4.3 Characterization

4.3.1 DC Characterization

The current and voltage characteristics were measured with an HP4145B parameter analyzer, with the chip in the dark, and then again illuminated with 1550 nm with an angled fiber through the grating couplers. The gate and drain voltages on each device were swept from 0 V to 1.5 V, while the drain current was measured. For high incident power measurements, an erbium doped fiber amplifier was inserted after the laser source to compensate for the insertion loss of the gratings and tapers.

Several devices with gate lengths from 0.5 μm to 2 μm and gate widths from 2 μm to 32 μm were measured. Measured transconductance was in the range 8-141 mS/mm and threshold voltages were in the range 0.5-2.1 V. The results for a 1 μm channel length and 8 μm channel width device are shown in Fig. 4.3. When operated in saturation, the device shows that for a bias current of 150 μA, a 51 μA photocurrent is produced when illuminated with 468 μW, seen in Fig. 4.3(a). The dark threshold voltage is 1.5 V from a linear extrapolation with a transconductance of 60 mS/mm. When 912 μW of 1550 nm light was incident on the device through the single mode waveguide, the extracted threshold was measured to be 1.285 V with the same transconductance, seen in Fig. 4.3(b).

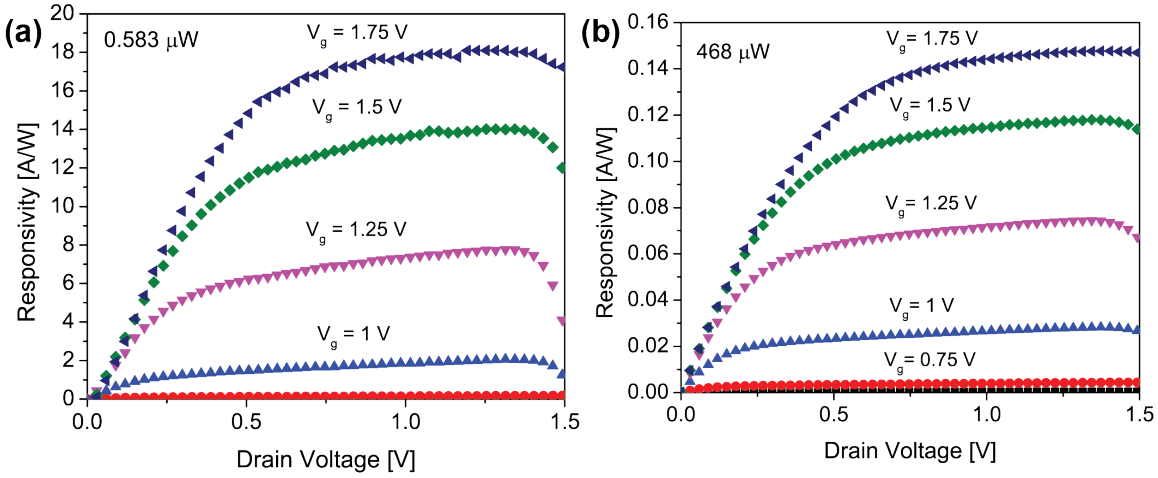

Figure 4.4: (a) Measured responsivity with 0.583 μW of incident power. (b) Measured responsivity with 468 μW of incident power.

The responsivity of the device can be extracted by dividing the photocurrent by the incident power, shown in Fig. 4.4, which shows responsivity as a function of gate and drain current. Much like transistor current, the photocurrent for a given input power is a function of the bias voltages. Higher gain can be obtained by increasing drain voltage, up to the point of saturation. Increasing the gate bias beyond the threshold voltage increases the photocurrent and responsivity. Because the device operates as a transistor with a photovoltage on the gate, the specific bias point of the transistor can be controlled by engineering the transistor itself, dictated by the specific output characteristic need.

The same transistor is then measured at several different optical input power levels. The photovoltage is extracted as a function of optical power, shown in Fig. 4.5. When fit with a linear function, a 60 mV/dec relationship is measured, agreeing with the previous theory. While this may appear to be a fundamental limitation to device performance, it should be noted that the intercept of the line is dictated by the internal diode leakage current. For this device, the extrapolated diode leakage current is 240 nA, which is an internal recombination current. Because there is a thick gate oxide, there is no measurable gate-source leakage current. The device can be made more sensitive, to yield a larger photovoltage for a given input power level, by shrinking the gate area, or by reducing the recombination rate in the gate through improvement in material quality or by surface treatments [86].

The log relationship between the input power and the induced photovoltage implies that how the voltage is then converted into a current is important. When operating the transistor in saturation, the drain current has a square dependence on applied gate voltage, and the amplification of small optical signals will be larger than that of large optical signals. This can be seen in Fig. 4.4 where the responsivity for 0.5 μW of incident light is a very large 18 A/W, while at 0.5 mW of incident light the responsivity is closer to 0.15 A/W. In the subthreshold region of operation, the drain current has an exponential dependence on applied

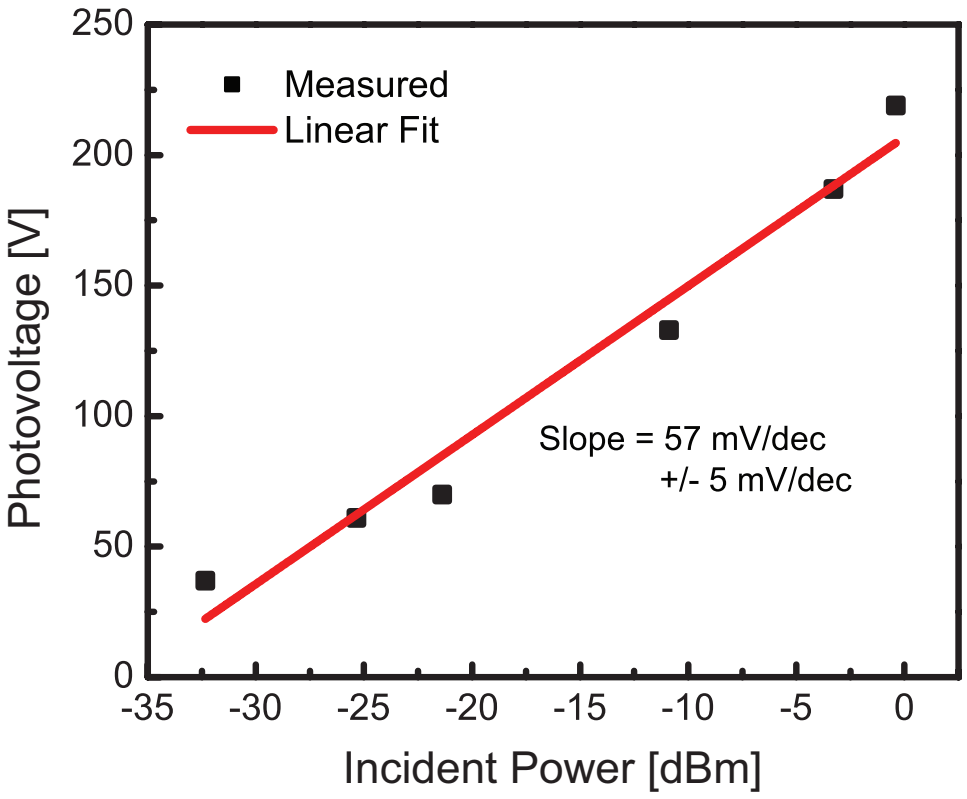

Figure 4.5: Extracted photovoltage as a function of optical power.

gate voltage, and there will be a linear dependence of drain current on optical power. However the overall output current is lower in subthreshold operation. When considering how to bias this device, one must consider device requirements such as linearity in optical response, output current, speed, and output impedance. Total photocurrent output in all regions of operation can be increased by decreasing the germanium diode area and by increasing the transistor transconductance.

4.3.2 RF Characterization

The high speed light response was measured with an Agilent 67 GHz network analyzer. The source and drain were contacted through a ground-signal-ground 50 GHz probe, with 1.2 V of drain bias applied through a bias-T. The drain current signal was put directly into port 2 of the analyzer. The gate was biased to 1.5 V with a DC probe. Port 1 of the analyzer drove a 20 GHz intensity modulator, which modulated the 1550 nm light input into the transistor. The frequency response was calibrated with a commercial 34 GHz photodiode. The frequency response measured with an input power of 912 μW shown in Fig. 4.6. The spectrum was fit with a single pole function, indicating a 3dB frequency of 2.5 GHz. A 2.5 GHz 3dB frequency was also measured at 1.37 mW of input power, which was the maximum

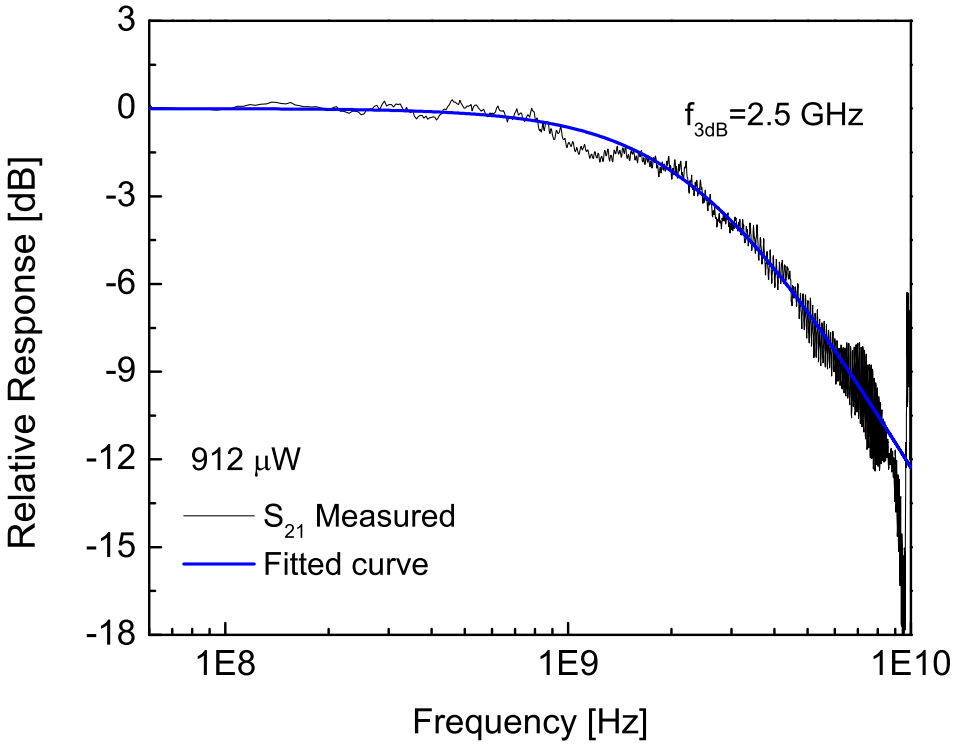

Figure 4.6: Relative response to optical excitation as a function of frequency.

available given the limitations of the laser source and erbium doped fiber amplifier, and also with 468 μW of input power. Below 468 μW of input power, the signal to noise ratio of the measurement became too low to accurately measure the 3dB frequency.

$$f_T = \frac{g_m}{2\pi\left(C_{ox} + C_{par}\right)} \tag{4.3}$$

where Cox is the gate oxide capacitance and Cpar is the parasitic gate capacitance.

We can calculate the cut-off frequency of the transistor using Eq. 4.3 [61], and use the measured transconductance, g_m= 0.46 mS, while assuming a gate capacitance from the patterned dimensions of 1x8 μm, and the measured 16 nm gate oxide, which gives a calculated gate oxide capacitance of 17.3 fF. With no parasitics this gives a 4.3 GHz cut-off frequency. In this particular device, the crystal seed was not trimmed from the gate, leaving over 25 μm of excess germanium, which should have about 5 fF of additional capacitance. Also because only a single deep source/drain implant was used, there is likely considerable gate-drain overlap capacitance in the device. Taking all of these into account, would then indicate that the measured 3dB frequency of the device is comparable to the natural cut-off frequency of the transistor. This indicates that by improving the doping profiles, and scaling the dimensions of the gate, the phototransistor should be able to operate at much higher speeds.

The other major component to the response time of this device is the relaxation time of the germanium gate. When the incident light is turned-off, the minority carriers present in the germanium must recombine to reduce the photovoltage present in the gate. This recombination lifetime is also directly correlated to the amount of photovoltage generated for a given amount of incidence light, since in steady state the generation and recombination rates balance. In this device, since the measured cut-off frequency is similar to the calculated theoretical cut-off for just the transistor, it is assumed the recombination lifetime in the germanium is not a limiting factor. However if the transistor is scaled to increase the bandwidth, a more detailed investigation of the germanium recombination lifetime will be needed to determine possible speed limitations from that effect.

4.4 Future Scaling

To investigate the potential benefits of scaling this device, we have run device simulations in Sentaurus. The specific process described previously was simulated to generate the device in the simulator, followed by electrical and optical simulations. To simulate optical excitation, a constant carrier generation model was used. First the simulation results were calibrated against the measured data. Constant field scaling was applied [61], where the gate oxide was thinned by the scale factor, the ion implant energy decreased by the scale factor, the gate length reduced by the scale factor, and the body implant dose increased by the scale factor. All other conditions remained the same, and the applied gate bias remained 1.5 V and the drain bias remained 1V. The per-unit gate width transconductance is extracted from each simulation, and from the geometrical dimensions, the gate oxide capacitance is calculated. Using Eq. 4.3, the cut-off frequency is calculated, and plotted against gate length in Fig. 4.7(a). The calculation assumes no additional parasitic capacitances on the gate. Additionally, the simulation of optical generation was calibrated with the measured data, and under the same conditions, the responsivity is plotted as a function of the scaled gate length in Fig. 4.7(b). In the simulation it is assumed that scaling does not affect the quantum efficiency of absorption, since a constant carrier generation model is used. In general this will not be true, since narrower gates will alter the confinement of light within the germanium, affecting the needed volume for absorption, and ultimately the total absorbed power. For scaling to very short gate lengths, additional optical engineering will be needed to ensure good absorption in small gate volumes.

It can be seen that by shrinking to 250 nm gate length, with 4 nm gate oxide, the cut-off frequency would reach 28 GHz. Following this same trend, a 100 nm gate length would be needed to reach 40 GHz, which is well within the capability of the state-of-the-area (22 nm node) technology. However one large assumption is that the carrier recombination time of the germanium gate diode is much faster than the transistor cut-off time. This time depends largely on recombination lifetime within the germanium, and will depend largely on material quality, and doping concentration in the germanium. As the gate shrinks in volume, surface recombination will become a bigger factor, and the dimensions of the gate will also affect

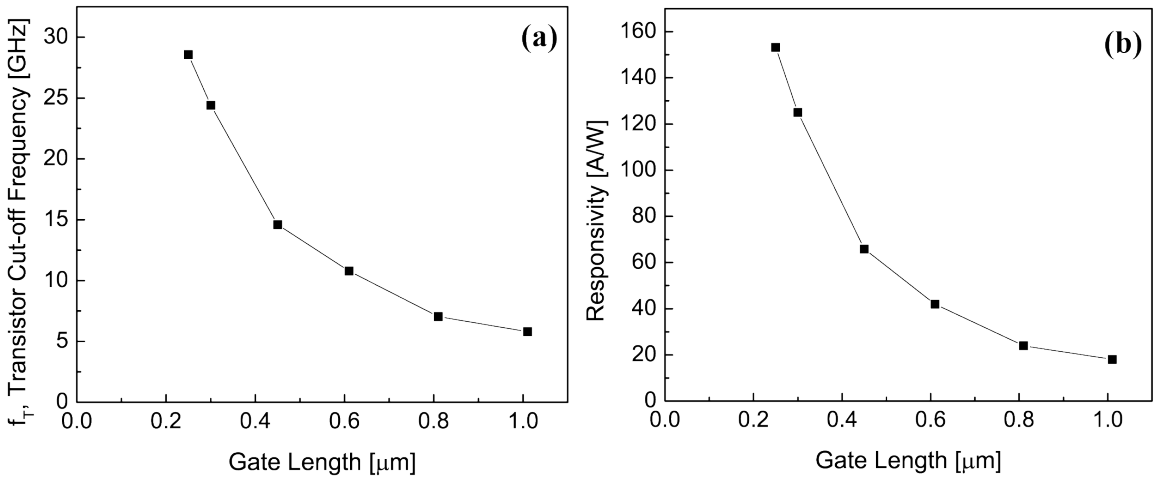

Figure 4.7: (a) Simulated cut-off frequency as a function of gate length with scaling. (b) Simulated responsivity as a function of gate length with scaling.

the germanium recombination lifetime.

In addition to a speed boost from shrinking the gate, the responsivity of the device should increase as well, which comes from two factors. One is that the induced photovoltage will increase with reduced gate area, since diode leakage current in the gate will reduce with the gate area. Additionally, a shorter gate will increase transconductance, meaning that for a given photovoltage, a larger drain current will be produced. When simulated, the 250 nm gate length device shows 150 A/W responsivity.

4.5 Linearizing Photovoltage through a bias network

As previously mentioned, one of the draw-backs of the photoMOSFET is the non-linearity of amplified current with respect to input optical power. This is largely due to the photovoltage induced in the gate because the photocurrent in the germanium is blocked by the gate oxide. We could imagine that if the gate were biased in such a way as to allow the photocurrent to flow through a bias resistor, the induced photovoltage on the gate, V_{photo}, would then be linear with respect to incident power, P_{inc}. Such a bias scheme is shown in Fig. 4.8, and is remarkably similar to that of a PIN-FET receiver[12]. Specifically, if we assume that the germanium gate has a quantum efficiency of absorption, η_{abs}, then the photocurrent, $i_{photo} = \eta_{abs} P_{inc}$, will drive the resistor, R_{bias}, such that $V_{photo} = R_{bias} i_{photo}$.

Such a bias scheme would require several alterations to the fundamental device layout and fabrication. It would be required to contact the germanium gate right at the gate oxide. Likely this would mean specifically doping the bottom layer of the germanium gate instead of relying on band-bending from the potential of the underlying silicon channel. This would ensure lateral uniformity of the gate potential at the gate oxide, and would allow for a

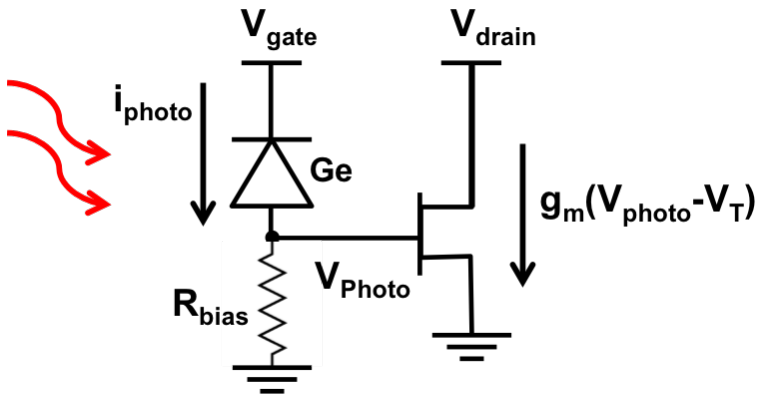

Figure 4.8: Proposed bias scheme to linearize the photovoltage applied to the gate of the photoMOSFET.

contact to that portion of the germanium gate to be placed laterally away from the MOSFET channel. Such a scheme would require additional masks and implantations.

The most critical aspect of this bias scheme will be the resistance of the bias resistor. Because the induced photovoltage will be proportional to the photocurrent, but there is a limited voltage range over which the phototransistor will be useful, between the threshold voltage, V_T and the supply voltage, V_{Gate}. In the example of the 1x8 μ m photoMOSFET previously shown, the threshold voltage is 1.5 V, and the supply voltage is 2 V. Thus if we assume a quantum efficiency of 100%, with a resistance of 1.5 MΩ, an incident power level of 1 μW would create a photovoltage of 1.5 V, and turn on the transistor. Any higher amount of power would saturate the gate at the supply voltage power and anything significantly lower would produce a photovoltage below threshold and the phototransistor would be essentially off. Thus the resistance must be chosen for the desired power level to be used, and the device would be very sensitive to swings in optical power. This is reflected in Fig. 4.9(a).

This biasing scheme will however have a significant impact on the speed of the phototransistor. Because the gate diode current goes through the resistor, the capacitance of the gate photodiode will be charged and discharged through that resistor. Since it was shown that in order switch the phototransistor from off to on with a low optical power, a large resistance is needed, it might be expected that this additional resistor will slow down the device. The RC 3dB frequency of this bias scheme is shown in Fig. 4.9(b), where $f_{3dB} = (2\pi C_{Ge} R_{bias})^{-1}$, where C_{Ge} is the capacitance of the germanium gate diode. The calculation is done assuming the previous lateral geometry of 1x8 μm, and an optimistic assumption that the junction width inside the gate is the total thickness of the gate itself, 350 nm. Using these numbers, the germanium diode capacitance is 3.2 fF. Thus for a 1.5 MΩ resistor, the 3dB frequency of the device will be 33 MHz. Using a smaller resistor will increase the speed, but at a cost of necessitating a higher amount of optical power to switch on and off the phototransistor. So while this bias scheme allows for a wider swing of the photovoltage and output current with respect to the incident optical power, it comes at a cost of a severe reduction in the

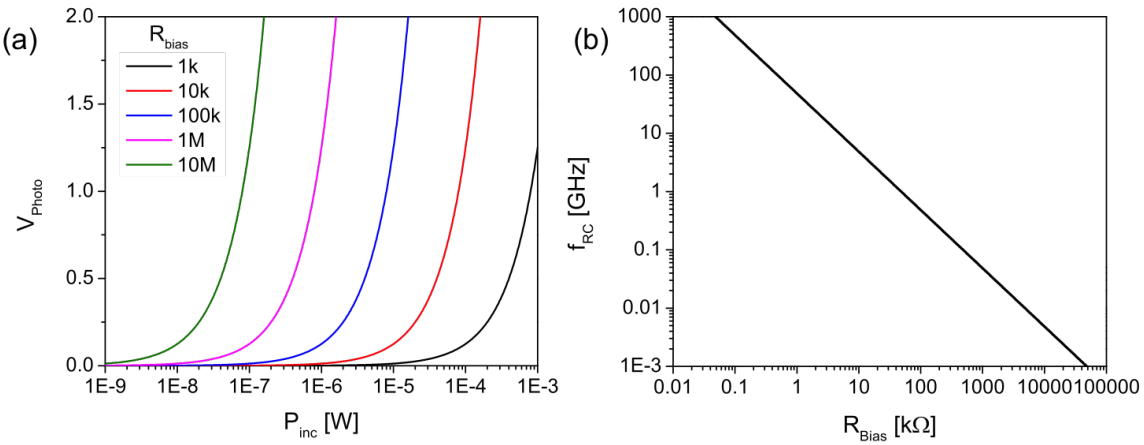

Figure 4.9: (a) Calculated photovoltage for various resistor values. (b) Calculated 3dB frequency as a function of resistance.

operating speed of the device.

4.6 Summary

This chapter has presented a germanium gated photoMOSFET, which demonstrates clear gain with a responsivity of up to 18 A/W, and a moderate bandwidth of up to 2.5 GHz. However there are some fundamental drawbacks with the photoMOSFET, such as the fact that current gain will be non-linear with respect to incident optical power, and that ultimately the response speed will be limited by the recombination lifetime in the germanium gate. For many of these reasons, our work then began to focus instead on a bipolar phototransistor design which is described in detail in the next chapter.

However because this is a MOSFET device, there are still interesting things which can be done with it, such as using CMOS scaling rules to reduce its capacitance, and increase its sensitivity and speed. Such scaling was done in simulations, which showed a dramatic improvement in these performance metrics. Moreover, even though the current gain is non-linear, one can imagine instead of using a single photoMOSFET, instead creating two photoMOSFETs, one nmos and the other pmos, to create an optically active inverter, which could then be tied directly to a more traditional digital amplifier or logic. Because of this flexibility, the photoMOSFET remains an interesting phototransistor device.

Chapter 5

Germanium Homojunction PhotoBJT

Although the photoMOSFET provides a familiar transistor structure and a clear scaling path almost identical to its electrical counterpart, there were some elements that are not as desirable as was shown in the previous chapter. Namely the intrinsic gain is non-linear, meaning the output current will not scale linearly with the input optical power. Additionally because the germanium gate is floating, there is not a clear current path for the photocurrent to dissipate, meaning there would need to be engineering of the material quality to optimize the recombination lifetime to make the device fast enough. One the other hand, a bipolar phototransistor has direct current gain of the base current, meaning the amplified output should be linear with respect to the input optical power, and base current is swept out through diffusion, meaning speed optimization is simply managed through control of geometry and doping.

Bipolar phototransistors are not new devices, and were first envisioned by Shockley[87], one of the original inventors of the transistor, where he simply described a bipolar transistor where instead of injection current into the base electrically, he would shine light on the device to generate the base current. More recently there have been attempts to scale a germanium photoBJT and build it on silicon to integrate it with silicon photonics[68]. They reported a gain bandwidth product of 7 GHz, which is a vast improvement over existing work, but in order for the photoBJT to be relevant for optical communications applications, this figure must be improved dramatically. In the following sections I will present the basic theory of operation for a photoBJT, some of our designs, initial results, and thoughts on how to build the ultimate photoBJT.

5.1 PhotoBJT Basic Theory of Operation

Figure 5.1 shows a basic NPN structure overlaid with a schematic band structure. The p-type base region forms a barrier to prevent electrons from flowing from the emitter to collector when a bias between the two is applied. In a traditional 3-terminal electrical BJT, a bias current is applied to the base, allowing electrons to flow into the collector and holes

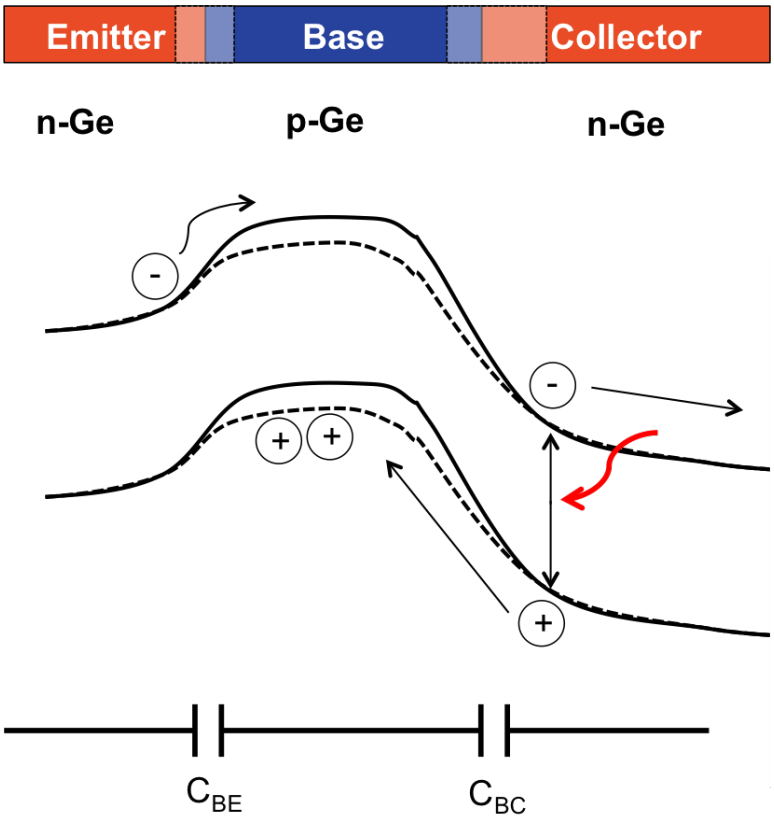

Figure 5.1: Example band diagram of an npn transistor.

become trapped in the base until they can diffuse over the barrier and through the emitter or recombine with electrons in the base. While this process occurs, the base barrier is lowered, and electrons more freely flow from emitter to collector. The overall current gain is then the ratio of the transit time of an electron from emitter to collector to the recombination time inside the base. This is commonly referred to as the β of the BJT. However for extremely compact BJTs where the base is very thin, generally the hole will diffuse from the base out through the emitter before recombining.

A similar process occurs for the phototransistor where now instead of an electrical bias, light is absorbed in the base-collector junction, and the photocarriers are swept away by the electric field present there, with electrons going to the collector and holes to the base. Again the holes become trapped in the base until they can diffuse over the barrier and through the emitter or recombine with electrons in the base. While holes are trapped in the base, the base barrier is lowered and electrons move more freely from emitter to collector. For a floating base phototransistor, there is not electrical contact to the base, and the total base modulation current is supplied through optical absorption. In a 3-terminal photoBJT, a combination of both electrical and optical bias can be applied to the base to get the desired output current in the collector. The relationship between total collector current and the

input base current, either electrical or optical photocurrent, can be approximated by

$$I_C = \beta I_B \tag{5.1}$$

β can be described in terms of the geometry and doping of the BJT since these are what affect the hole and electron transit times. Here it is assumed that the hole diffusion length is much longer than the emitter width, which should be the case for the compact BJTs examined here.

$$\beta = \frac{N_E W_E D_{nB}}{N_B W_B D_{pE}} \tag{5.2}$$

where N_E and N_B are respectively the doping concentration in the emitter and base respectively, W_E is the emitter width or depth, W_B is the quasi-neutral base width or distance from base-collector depletion edge to base-emitter depletion edge, and finally D_{nB} and D_{pE} are the diffusion constants for electrons in the base and holes in the emitter respectively [72].

Finally, we can describe the overall speed of the transistor[72]

$$\frac{1}{2\pi f_T} = \tau_F + \frac{kT}{qI_C}\left(C_{BE} + C_{BC}\right) \tag{5.3}$$

where f_T is the cut-off frequency of the transistor, where the current gain becomes 1, or depending on the biasing scheme, can also be thought of as the gain-bandwidth product of the transistor. τ_F is the forward transit time, or the total time it takes for an electron to travel from emitter to collector. The second term can be thought of as the RC delay component of the transistor, where the capacitances are the two p-n junctions, and the resistance is given in terms of the transconductance, which is modulated by I_C.

From these equations a few generalizations about the design of the transistor and phototransistor can be made. First is that since the emitter and base quasi-neutral region widths, W_E and W_B, will largely be similar and to an order of magnitude approximation the diffusion coefficients, D_{nB} and D_{pE}, will be as well, β can be approximated by simply the ratio between emitter and base doping. Additionally, in order to make the device fast, we must make the capacitances as small as possible, so junction areas should be kept a small as possible, and the transit times should be kept short by making the junctions overall as thin as possible. All of these combine to create challenges associated with fabricating such a small and thin device, but the further sections will discuss our potential solutions.

5.2 3-Terminal PhotoBJT Design

In creating our initial 3-terminal photoBJT design, we wanted to create an extremely high gain-bandwidth device, which meant that both the npn junctions needed to be very thin to make transit time as short as possible, and that the junction areas needed to be as small as possible in order to reduce base-emitter and base-collector capacitances. A basic schematic of

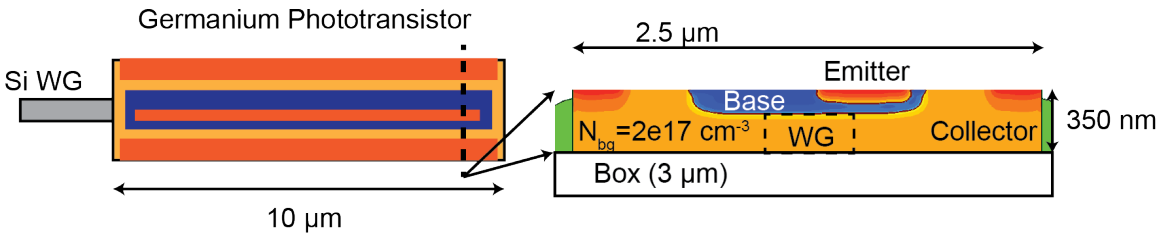

Figure 5.2: Schematic of a 3-terminal photoBJT butt-coupled to a silicon waveguide with lateral npn arrangement.

the device is shown in Fig. 5.2. To take advantage of the existing RMG process, a lateral NPN doping scheme was used. Because the device is sitting on oxide, we cannot have a bottom contact, so the emitter, base, and collector contacts must all be made laterally. In the future, the RMG process could be modified to allow for a bottom silicon sub-collector contact for a more traditional vertical npn by using micro-bonding after RMG[57] for example.

The lateral dimensions were chosen to minimize capacitance, but at the same time, we are constrained by the minimum resolution of the DUV stepper in our nanfabrication lab. In order to process on 6" wafers with the DUV stepper, the minimum feature size was chosen to be 300 nm, which in this case was the separation in the metal layer between emitter, base, and collector contacts. Included were 100 nm overlays, since the layer-to-layer alignment tolerance of the stepper is +/- 50 nm. Thus the emitter was chosen to be 500 nm wide, and go along the length of the phototransistor. In order to have the majority of the current flow through the implantation defined emitter/base/collector junction, the base is designed to have at least 200 nm of spacing on either side of the emitter, making the shortest current path vertical through the implanted junction. With enough space to include a metal contact to the base, this makes the base region 1.2 μm wide, and it too goes along the length of the phototransistor. Finally the heavily doped collectors were placed along the outside of the device, and were also designed to be at least 300 nm wide, both for critical lithography considerations and to improve contact resistance. These were placed as closely as possible to the base and emitter in order to reduce transit time. This makes the total width of the device 2.5 μm which is close to the maximum width that can reliably be made with germanium RMG. Finally, the base is placed in the center of the phototransistor such that when the silicon waveguide is butt-coupled to the center of the germanium, the majority of the light will propagate into the center of the device and absorb along the base/collector junction.

The thickness and length were largely determined by trading-off optical absorption simulations done with FDTD with electrical simulations done in Sentaurus. Optical FDTD simulations were performed for various thicknesses and lengths of germanium when butt-coupled and it was found that the absorption was largely independent of the thickness when simulated with a thickness range of 200 to 600 nm. However the primary factor in total absorption was length, as might be expected, and largely the absorption followed the Beer-Lambert law of absorption given the germanium absorption coefficient. This means that

for example, a 10 μm long germanium device will absorb about 67% of the light incident from the waveguide, since the absorption coefficient for germanium at 1550 nm is about 1000 cm^{-1}[74]. However, because the emitter/base/collector junctions are defined vertically through ion implantation, there are severe trade-offs with regards to electrical performance and germanium thickness. Specifically if the germanium is too thin, the collector region underneath the base becomes very thin, and the speed is limited by current crowding. Making the collector region thicker, allows more current to flow, and hence the f_T increases towards its transit limited maximum. However if the germanium layer is too thick, then photocarriers generated in the collector have much farther to travel, and the f_T measured when exciting the photoBJT with an optical signal will be reduced due to the delay in photocarrier collection. It was found that 350 nm of germanium thickness for the given emitter and base implantation profile afforded a good trade-off between current crowding and fast photocarrier collection.

In general the doping profile had a few goals in mind. First we wanted a moderate gain of about 10 in order that the 3dB bandwidth could be in the several GHz range. This meant the emitter doping should be about 10 times more than the base doping. Secondly, we wanted the base transit time to be fast, on the order of 1 ps, which corresponds to a base thickness of about 100 nm. Trying for a thinner base would certain yield a faster device, but due to concerns about diffusion of the phosphorus for the emitter[83] during activation, 100 nm was deemed sufficient. Additionally, the emitter was designed to be about 60 nm thick, since it was seen in the p-i-n device that when etching into the oxide to contact the germanium about 30-40 nm of germanium was removed during the over-etch. The emitter should be thin, but we did not want it to be so thin that it would get etched through during processing.

Table 5.1: Implantation doses and energies for the phosphorus emitter implant.

Implant	Energy [keV]	Dose [cm^{-2}]	Angle
1	25	10^{14}	7°
2	15	$5(10^{13})$	7°
3	5	$2.5(10^{13})$	7°

Table 5.2: Implantation doses and energies for the boron base implant.

Implant	Energy [keV]	Dose [cm^{-2}]	Angle
1	25	$1.5(10^{13})$	7°
2	15	$7.5(10^{12})$	7°
3	5	$5.6(10^{12})$	7°
4	2.5	$3.75(10^{12})$	7°

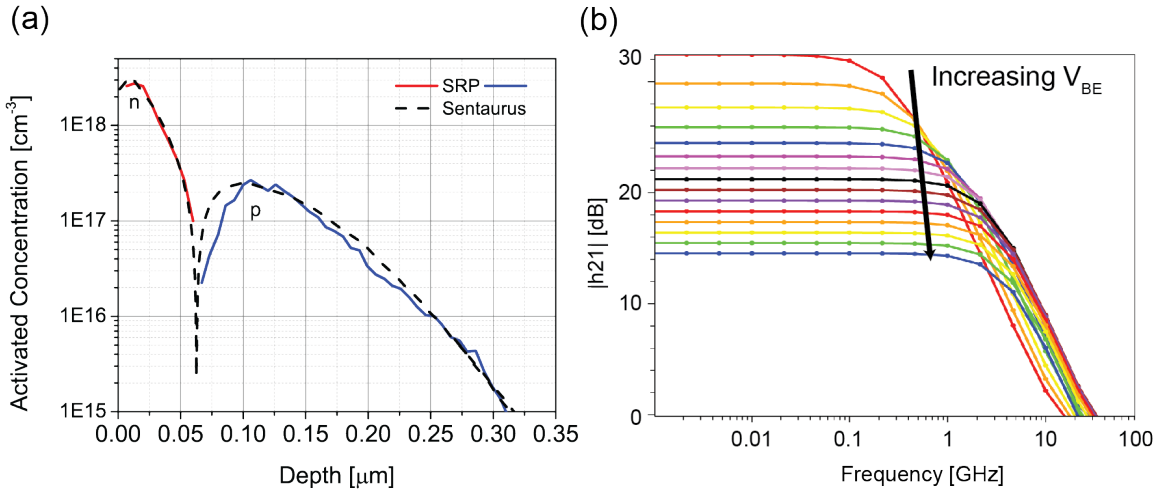

Figure 5.3: (a) Shows SRP results from the optimal base/emitter implant profile, along with the simulated profile from Sentaurus. (b) Shows simulated RF performance of the 3-terminal photoBJT with the calibrated implantation profile and a background concentration of $2(10^{17})$ cm^{-3}.

To calibrate the doping profile, several short-loop experiments were done. First the emitter and base implants were simulated, and when an appropriate profiles were found, we implanted these into bulk germanium wafers, which were then thermally processed identically to the existing process to recreate the activation conditions. Although simulation of the implant energy and depth is quite accurate, there is less certainty about the amount of activation for the dopants, and so multiple doses and combinations of doses were chosen and applied to multiple samples to find the optimal combination. In order to extract the activated doping profile, after the thermal activation of the implants, the samples were sent to a vendor for scanning resistance probing (SRP) which grinds the sample at a shallow angle and measured the 4-point probe resistance along the grade. Doping type can also be determined, and the resulting SRP doping profile can be seen in Fig. 5.3(a). Bulk germanium wafers were used for this short-loop experiment because SRP profiling generally needs a large area sample, whereas the RMG technique can only be used to produce strips of germanium 3 μm wide. Tables 5.1 and 5.2 show the final optimized implantation scheme used in the fabrication of the photoBJTs.

The implantation profile was recreated in Sentaurus simulations, and the device as shown was simulated to extract the electrical and optical performance. The electrical RF performance can be seen in Fig. 5.3(b), which shows gain vs frequency for several bias conditions on the base, with a collector emitter bias of 0.8 V. It can be seen the DC gain is about 10 (20 dB), which is consistent with expectations from the doping profile. However the 3dB frequency is only about 3-4 GHz, indicating a peak f_T of 32 GHz. This is likely limited by the low background collector concentration and the thickness of the base. However it is a

very good starting point for a photoBJT device, and when simulated with an optical signal, a similar f_T is obtained, indicating good photocarrier collection as well.

To illustrate the impact of the background doping, several simulations were done, which varied the background concentration, and also simulated the effects of having twice the calibrated dose for the base and emitter implants, which can be seen in Fig. 5.4. What can be seen is that as the background collector concentration is increased, the peak f_T increases as well, until it reaches a peak where the collector concentration is approaching the concentration of the base, and thus the gain begins to decrease due to punch-through. This onset of punchthrough at higher collector concentrations can easily be seen by the increase in emitter-collector leakage current, which is the current flowing through the collector when collector is biased, but no current is flowing into the base. At lower collector concentrations the peak f_T is slower both because the base is thicker, but also because of the lower collector concentration, there is an earlier onset of the Kirk effect[61], which limits the f_T beyond a certain current. Because the overall doping level of the emitter and base are still a bit low, a second set of curves show the effect of doubling their emitter and base implant doses. This allows for higher collector concentrations, reducing the Kirk effect, and ultimately allowing for a higher f_T. Because of the sensitivity of the device performance to small changes in the collector background concentration, several wafers were prepared with different background implantation doses.

5.3 3-Terminal and Floating Base PhotoBJT Fabrication

The fabrication process is largely similar to the previously described processes with a few modifications. We start with a 6" SOI substrate with a 3 μm Box and 220 nm silicon layer. First the pre-alignment marks for the DUV stepper are exposed and etched using a silicon RIE etch tool. Next the waveguide layer is exposed using thin (420 nm) DUV resist, hard baked, and etched fully down to the Box layer using the silicon RIE tool. In order to ensure the dimensions on any of the critical feature size layers, a wafer is first exposed with a matrix of exposure doses, hard baked, and examined under SEM to measure the feature sizes. After the waveguide layer has been etched, the grating coupler layer is exposed, and etched to a depth of 60 nm with RIE. The etch time is calibrated using a plain silicon wafer, and the etch depth is checked with a confocal microscope.

Next the wafers are pirhana cleaned, and placed into the oxidation furnace, where 20 nm of dry thermal oxide is grown at 1000 °C for 14 minutes with a 5 minute anneal in nitrogen. Next the seeding layer is exposed, and the wafers without a hard bake, are placed in dilute buffered HF to removed the oxide from the region used as the crystal seed for RMG. After stripping the resist in PRS-3000 so as to not oxidize the substrate further, the wafers receive a brief HF dip and are immediately placed in the germanium LPCVD furnace, where 350 nm of germanium is deposited. Upon removal from the germanium furnace, 20 nm of LTO

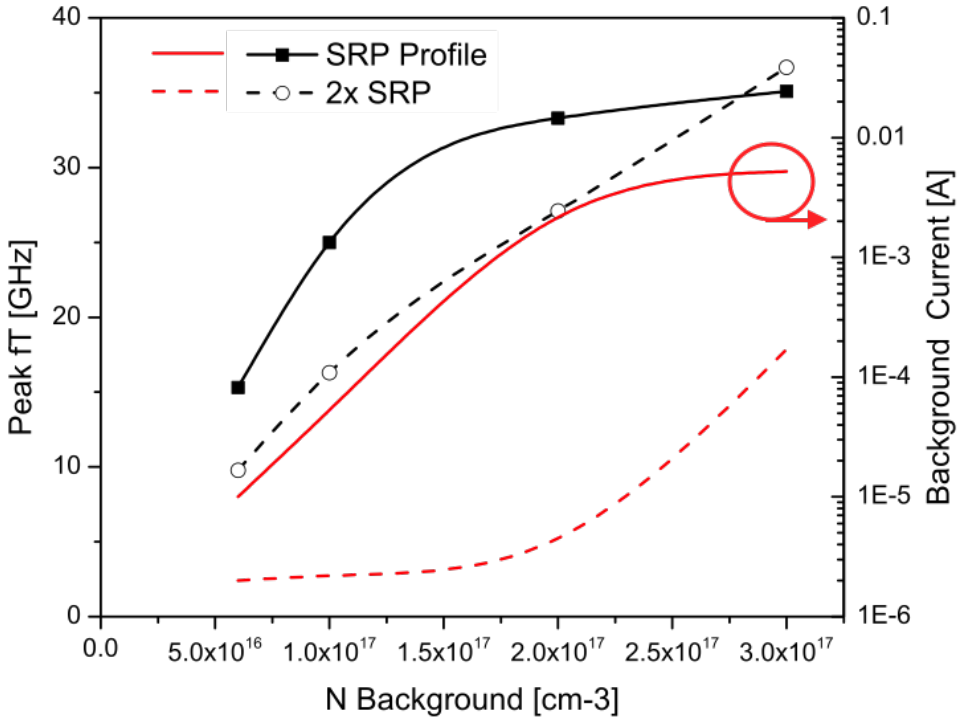

Figure 5.4: Summary of simulated f_T for the optimized implant profile from SRP, with various background concentrations. Also shown is the collector-emitter leakage current resulting from base punch-through. A second set of curves show the same simulations if the doses for the base/emitter implant are doubled.

is deposited on top to provide a stiction layer for the photoresist is subsequent lithography steps. The germanium mask is then exposed and hard baked, and the germanium layer is etched in the silicon RIE etcher down to the Box layer. To remove the remaining strands of germanium along the waveguides, instead of the traditional germanium clean mask, a simple HF dip is used, which undercuts the oxide underneath these small residual strands and they are washed away. This was found to be as good as selectively etching the remaining germanium with peroxide.

The backside germanium is removed with 50 °C peroxide while the front side of the wafer is protected with resist. The wafers are then shipped out for the background implant, which uses a 100 keV phosphorus implant, and various doses to alter the background concentration across the wafers. The wafers are then placed in the LTO furnace, and 400 nm of LTO is deposited on the wafers. They are then placed in the rapid thermal annealer for the RMG process, which also activates and uniformly distributes the background phosphorus doping.

The wafers then are placed in an oxide RIE etcher and receiver a blanket plasma etch to etch all but 50 nm of the oxide on top of the germanium. The remaining oxide is removed with dilute HF, and a thin, 5nm, LTO is re-deposited on top for resist stiction. The base

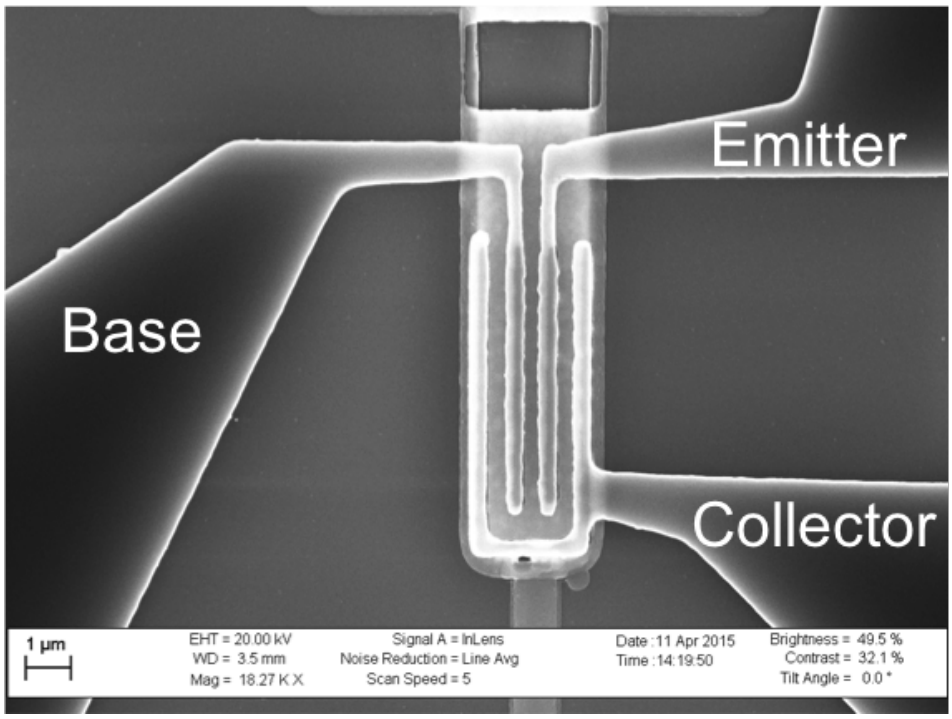

Figure 5.5: SEM image of a competed 3-terminal photoBJT device.

implant mask is then exposed on the wafers, and after hard baking they are shipped out for boron implantation using the implant profile in Tab. 5.2. After the resist is stripped, the emitter implant mask is exposed and hard baked, and the wafers are again shipped out for phosphorus implantation using the profile in Tab. 5.1. After stripping the resist, a 250 nm layer of LTO is deposited, and the implants are activated in a rapid thermal annealer at 500 °C for 30 seconds.

The contact opening mask is exposed, hard baked, and the contact holes are etched in the oxide RIE. The wafers are then placed in the sputtering tool, where a short 1 minute sputter etche is performed to remove any surface oxide, then 50 nm of Ti and 450 nm of $Al_{98\%}Si_{2\%}$ are deposited. Finally the metal contact mask is exposed, and the metal layer is etched in the metal RIE etcher, overetching to fully remove any metal from the sides of waveguides.

An representative SEM of the completed 3-terminal photoBJT is shown in Fig. 5.5. The fabrication run using that design and process was just recently completed and the initial results indicated that the emitter had direct contact with the collector through the base, but diagnostic measurements were still being performed at the time of this writing. This should indicate what about the process, whether the implant energies or doses, or the activation temperatures need to be modified to create working 3-terminal photoBJTs in the future. To first demonstrate a preliminary device, a simpler fabrication process was first performed, which was identical to that used in the fabrication of the PIN photodiodes in Ch. 3, to create

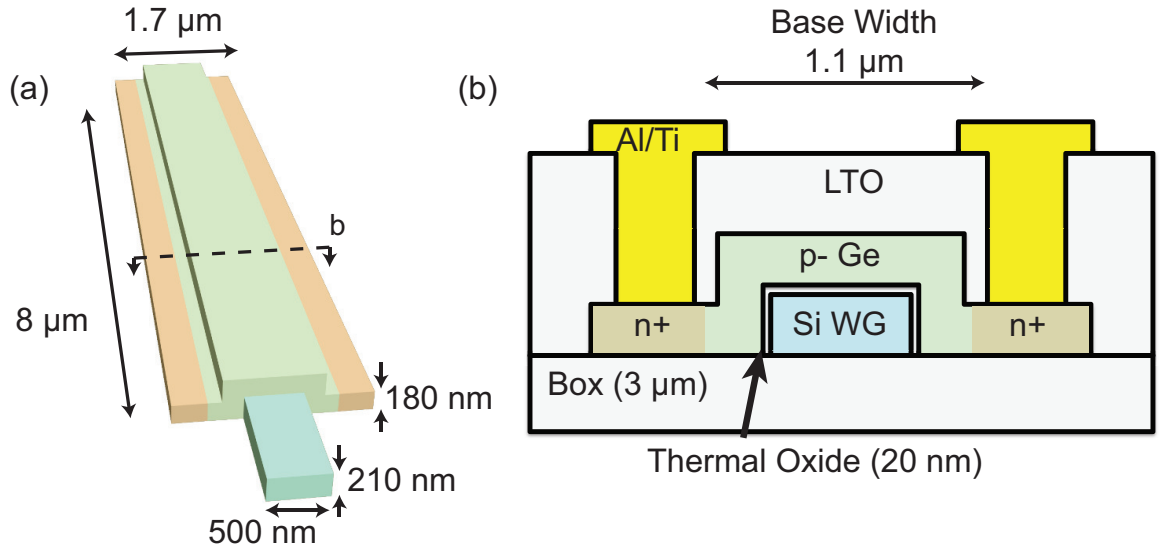

Figure 5.6: (a) Schematic of the 2-terminal floating base photoBJT, geometrically very similar to the wrap-around photodiode. (b) Schematic cross section of the floating base photoBJT, showing n+ emitter and collector on the sides and a large p- base in the center.

floating base NPN photoBJTs, which were characterized and are described in the following section.

5.4 Preliminary Floating Base PhotoBJT Devices and Results

The preliminary photoBJT devices which were measured were of a nearly identical design to the the PIN photodiodes described in Ch. 3. They are a much simplified version of the 3-terminal BJT described in the previous section. The schematic of these preliminary devices is shown in Fig. 5.6. Much like the wrap-around PIN photodiodes in Ch. 3, these are wrap-around NPN phototransistors. Here the doping placement is identical to the wrap-around PIN except both sides are doped heavily n-type, and the middle portion will be lightly p-type, due to the natural p-type nature of germanium. It was shown in the investigation of the PIN device that the intrinsic germanium produced with RMG is approximately 10^{16} cm^{-3} p-type. Thus there is no separate background doping or explicit base doping for these BJTs as was described in the previous section. This is simply to simplify the process, using one identical to that used to create the PIN devices, and initially explore the characteristics of the NPN phototransistor. Additionally these photoBJTs have only two contacts, and thus the base is left floating. This means that all base current must be supplied by photocurrent from absorbed incident light.

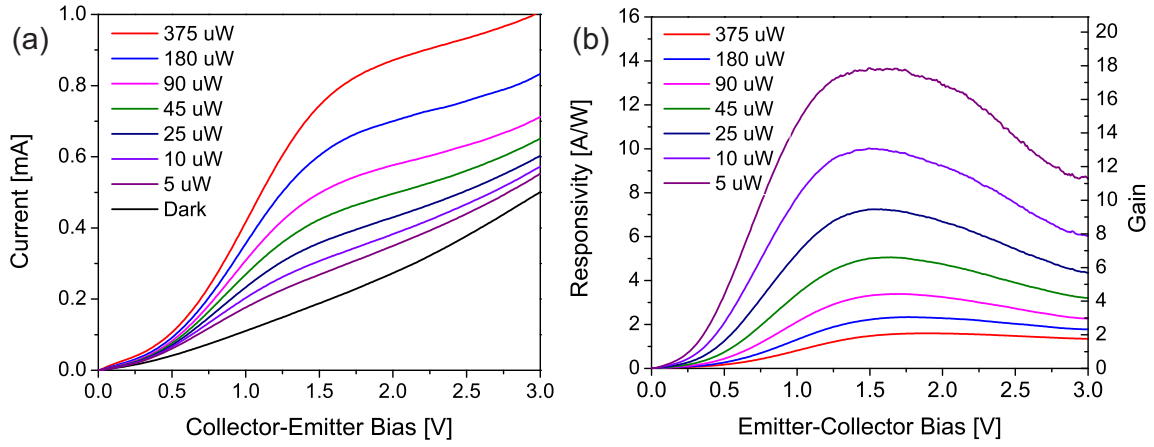

Figure 5.7: (a) Shows the I-V characteristics for a 1 x 8 μm floating base photoBJT with various input optical powers. (b) Shows the measured responsivity and gain of the photoBJT when normalized against a geometrically similar p-i-n photodiode.

Several different length devices were made, but the primary device measured here is one with 8 μm of length along the waveguide. The low-doped base region is 1.1 μm in width, and the germanium itself is 180 nm thick. Thus the quantum efficiency of absorption for this NPN phototransistor should be identical to that of the 8 μm long wrap-around photodiode in Ch. 3, and we can use the measured responsivity of that photodiode to estimate the actual photocurrent gain in the NPN, that is an intrinsic responsivity of 0.75 A/W at 1550 nm. The I-V characteristics of the photoBJT were done by applying a voltage across the emitter and collector using a parameter analyzer, and then using the grating couplers to input different amounts of optical power, much like with the photodiodes.

From the I-V curves seen in Fig. 5.7(a), we can see several key characteristics of this preliminary photoBJT device. First, we see a clear increase in photocurrent with increasing optical power, which is to be expected. On all of the curves with optical power input, we see a rapid rise in current with increasing voltage until a specific voltage is reached, around 1 V, where the current then saturates and levels off. This is very typical of a BJT and is nearly identical to a text book I_C vs V_{CE} family of curve with increasing I_B [61]. What can be seen is at lower voltages, the device is in the saturation mode, while above the critical voltage, the device moves into forward active mode. The voltage increases as increasing amount of base bias current is applied, which we also clearly see in the figure.

Another characteristic seen is that even in forward active mode, there is some noticeable slope to the I-V curves. In a BJT this slope is the differential output resistance, and for a well designed device the resistance should be as large as possible. Having a low output resistance implies a small Early voltage, or likewise a strong base-width modulation effect [61]. Base-width modulation is the effect where the base-collector depletion region width is strongly affected by the base-collector voltage, which would be very common for a base with

low doping. This makes sense for these phototransistors, since the assumed base doping level is about 10^{16} cm^{-3} p-type. As the optical power increases, the base-width modulation effect decreases, likely because with more carriers generated in the base with higher optical power, the net carrier concentration inside the base increases, requiring more voltage to deplete it.

Finally we can see that as the voltage becomes larger, near 3 V, the current begins to increase much more rapidly. This is likely the onset of breakdown or punch-through, where the base-collector depletion region has expanded to the point that it is beginning to overlap with the base-emitter depletion region, leaving little to no undepleted base region [61]. Again because the effective carrier concentration is higher with higher optical power, punch-through can be seen to occur at higher voltages when the optical power is higher. Overall the dark current for this device is quite high, much higher than the photodiode, because there is a strong amount of punch-through even at low voltages, which allows current to easily leak from emitter to collector.

All of these effects can easily be seen when the dark current is subtracted, and the optical power normalized in the responsivity and gain plot in Fig. 5.7(b). Here the responsivity increases and then plateaus when transistioning from saturation to active mode, then begins to drop as a result of the Early effect and beginning of punch-through. The gain reduction with increasing voltage is much less for higher optical powers because the base-width modulation and punch-through effects are less severe. As mentioned earlier, it is assumed the intrinsic responsivity is 0.75 A/W, taken from the photodiode data, and thus the gain is simply the measured responsivity divided by this intrinsic responsivity.

It can be seen the gain and responsivity are quite high for low optical power, up to 14 A/W or a gain of 18, and then decrease with increasing optical power. While this may appear similar to the photoMOSFET response, the cause is quite different. In this case, because the doping level in the base is so low, high injection effects appear with very little base current, or in this case optical power[61]. These high injection effects are more apparent in the RF response of the device, shown in Fig. 5.8.

The S_{21} spectra were obtained using a PNA just as in previous chapters and were measured for optical power levels identical to those used in measuring the DC I-V characteristics. In this case, however, the DC level of the S_{21} curve was offset by the gain calculated previously. This results in the h_{21} spectra seen in Fig. 5.8(a). As was apparent from the I-V curves, as optical power increases, the DC gain decreases. However because in an ideal BJT, the gain-bandwidth product should remain largely fixed, as the optical power decreases and the gain decreases, the bandwidth also increases. To describe the gain-bandwidth product, we can also use f_T, which is the unity gain frequency, or where the h_{21} curve crosses the 0 dB point[61]. We can see that the peak f_T occurs for the lowest optical power actually, and then slowly begins to decrease. This is more apparent in Fig. 5.8(b) which plots f_T as a function of optical power and V_{CE} bias. This shows what was described earlier, which is that even with only 5 μW of optical power, the floating base photoBJT is already suffering from high-injection effects, which limit the f_T, most likely through the Kirk effect[61]. The Kirk effect is more acute for BJTs with low doping levels, since it is caused when the base-collector depletion region extends through the collector to the more heavily doped sub-collector, and

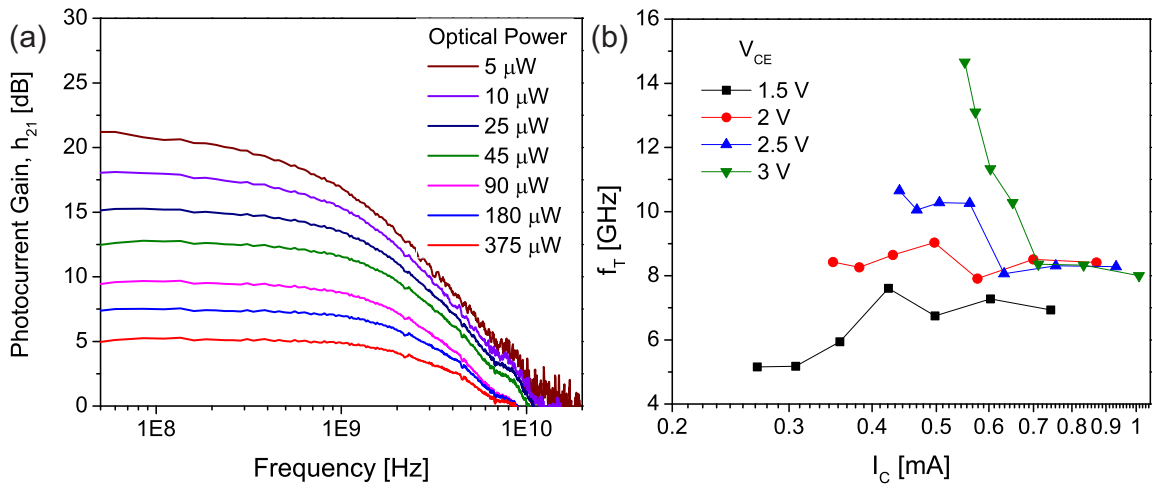

Figure 5.8: (a) h_{21} spectra for the 1x8 μm photoBJT at various optical power levels with $V_{CE} = 3$ V. (b) Extracted f_T for various optical power levels and V_{CE}.

the electric field becomes very high at this highly doped region, rather than being uniform across the entire base-collector depletion region.

Thus it can be seen the f_T for the preliminary photoBJT is as high as over 14 GHz, but at low optical power levels, high dark leakage, since it is operating near punch-through, and with low 3dB bandwidth. This means for the photoBJT to be usable, the following stage must include some equalization to compensate for the low 3dB bandwidth to allow for a flat response up to the desired bit rate. Additionally, if one were to desire a reasonable amount of gain at a moderate bit rate, such at 10 GB/s or even 25 GB/s, the f_T must be increased event further, likely well above 100 GHz, as this would make the gain-bandwidth product more competitive with state of the art avalanche photodiodes[4]. The following section will discuss how it is possible to scale the current photoBJT, and improvements which can be made to its design to get an even higher f_T.

5.5 photoBJT Scaling Considerations

The photoBJT can be scaled to higher gain-bandwidth products much in the same way that the photoMOSFET can be scaled. Eq. 5.3 shows two primary components which limit the speed of the transistor. One is the overall capacitance of the base-collector and base-emitter junctions. This can be scaled by decreasing the junction lateral areas and also decreasing the overall junction thicknesses. The base-emitter junction, with both sides generally being more heavily doped than the base-collector junction, will in general have a thinner depletion region, and thus higher capacitance. So it is most important in this consideration to scale the area of the emitter itself.

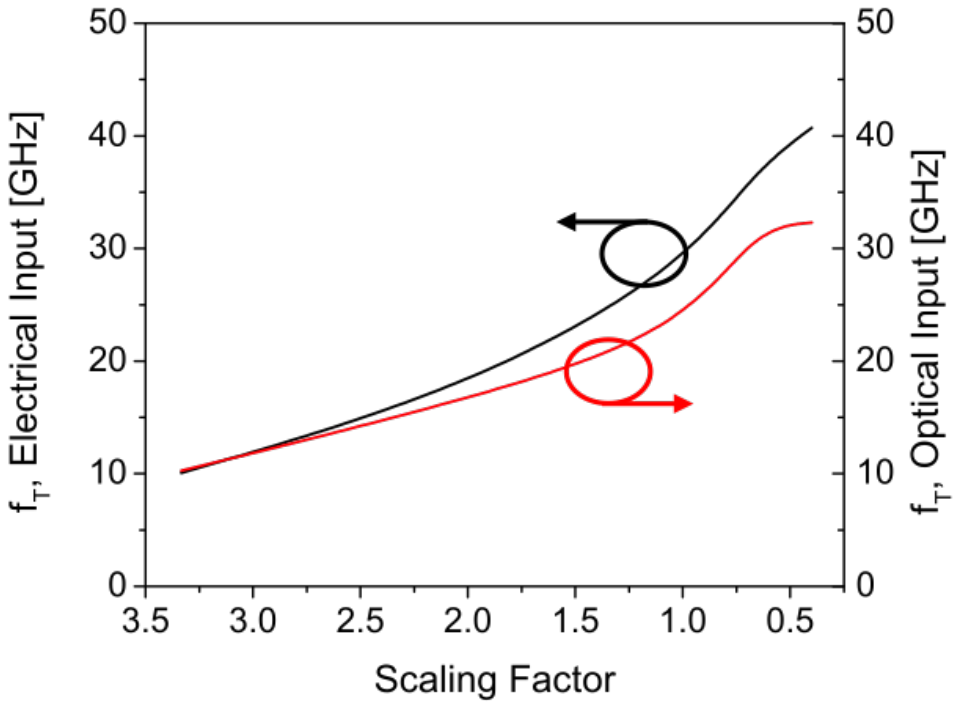

Figure 5.9: Effects of the perceived f_T with both electrical and optical input signals when scaling the photoBJT.

The second factor is the overall transit time of electrons from the emitter to the collector. This transit time is decreased by shrinking the thickness of the base region between the emitter and collector. To prevent punch-through of the base, when thinning it, the base doping should be increased as well. Since increasing the base doping will decrease the overall gain, the emitter doping should be increase proportionately as well. Finally because pushing to higher f_T will also require higher currents, as seen from the I_C term in the equation, to prevent the Kirk effect at high biases, the collector doping level should increase as well. Thus overall, to scale the photoBJT, the emitter area should decrease, the base region thickness should be decreases, and the overall doping level should increase for the emitter, base, and collector.

The results of this scaling with the original photoBJT structure shown in Fig. 5.2 are shown in Fig 5.9. In this calculation, the emitter area is kept constant, because we have already scaled the emitter area to as small as can be reliably resolved with out existing DUV lithography available in the nanolab. The scaling factor of 1.0 assumed an emitter/base implant profile identical to that shown in the previous section with emitter doping of $6(10^{18})$ cm^{-3} and base doping of $6(10^{17})$ cm^{-3}, with a uniform n-type background concentration of $3(10^{17})$ cm^{-3}. The simulated f_T for this configuration is 28 GHz. To approximate the effects of scaling, we then scale only the uniform background doping, since increasing this will shorten the collector region, and also thin the base region. Thus the scale factor refers

to the original background doping relative to simulated background doping. Thus a scale factor of 0.5 has twice the background doping and a much thinner base due to the p-type implant profile.

What can be seen is that as the background doping is increased, and the base thinned, the f_T rises dramatically, well over 40 GHz with a scale factor of 0.4. However, because the geometry is not altered, but the collector doping has increased, the f_T when simulated using an optical input signal increases more slowly and plateaus at about 30 GHz. The reason for this is that because the collector doping as increased, for a given voltage, the base-collector depletion width is now narrower, so a larger percentage of the photocarriers which are generated must now diffuse into the base rather than move by drift in the high electric field depletion region. Additionally, because the doping level is higher, their mobility has decreased as well. Overall this means that although the transistor itself is now faster with scaling, the overall phototransistor is not any faster because the photocarriers are collected more slowly. To overcome this, the thickness of the photoBJT could be decreased or the overall width as well, but at the expense decreased absorption or dimmensions which are too small to obtain with the DUV lithography equipment available in the nanolab.

As an alternative which is described in the next section, the transistor gain region can be effectively de-coupled from the absorption region, creating both a fast transistor with fast photocarrier collection, to create an overall fast photoBJT.

5.6 De-coupling Gain from Absorption for Ultimate Gain-Bandwidth

As has been done with avalanche photodiodes, where a separate absorption and multiplication region are formed to increase the gain-bandwidth product while maintaining quantum efficiency[4], so can the gain and absorption region of a photoBJT be separated and co-designed. Fig. 5.10 shows an example of this design, with a purely vertical junction structure in the center of the photoBJT, with very thin, 30 nm, base which is also highly doped, a narrow, 100 nm, emitter, and a highly doped collector leading to an underlying silicon sub-collector. While in general there is nothing extraordinary about this design, since it appears very similar in structure and dimensions to advance SiGe BiCMOS BJTs[61], it is what surrounds the transistor region that is more interesting.

By having a region surrounding the central transistor area which has only the base and collector regions, we can have a large absorption area with fast photocarrier collection. Specifically the example shows a thick, 300 nm, region with a highly doped base, and highly doped sub-collector, but intrinsic germanium between them. This region looks almost identical to that of a vertical PIN, and since in active region operation, the base-collector is reverse biased, photocarriers generated in this region will rapidly transit to either the base of collector. Once in the base, where the holes are majority carriers, the base barrier in the transistor portion of the device will lower, giving transistor gain. Because the primary

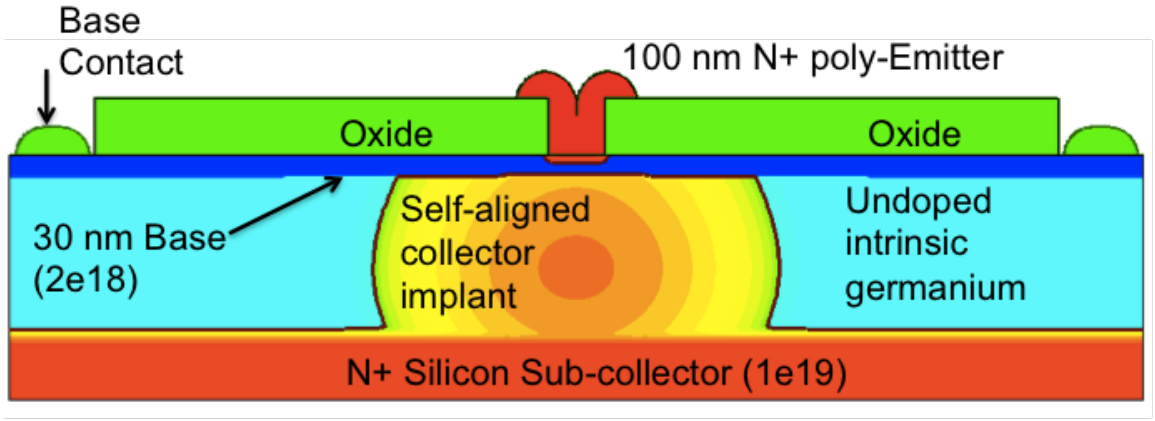

Figure 5.10: Example photoBJT which has been highly scaled and exhibits a separate gain and absorption region.

current path for electrons to travel to collector is through the vertical emitter/base/collector region of the photoBJT which has been shrunk vertically, the electron transit time will be very fast, leading to a high f_T.

This can be summarized with a simple example simulation with this example structure, which is shown in Fig. 5.11. The photoBJT is first simulated with an electrical input signal, to estimate the f_T of the device, which as has been shown previously will have some dependence on the bias applied to the base. At the optimal bias, the f_T peaks at about 120 GHz, with a gain of 6 and a 3dB bandwidth of 20 GHz. When a similar electrical bias is applied to the base, an optical signal is simulated by having uniform elctron-hole generation across the device for 1 ps, creating an impulse, which can be converted to a frequency response using an FFT of the collector current time signal. The extracted gain-bandwidth product is found to be 108 GHz, showing that while the gain is still 6, the 3dB bandwidth only dropped to 18 GHz with an optical signal. While this is simply an example structure, and is largely unoptimized for peak performance as a photoBJT, it does illustrate how the gain-bandwidth product for a photoBJT can be made very fast while keeping dimensions of the structure large enough for efficiency optical absorption.

Finally, we can begin to imagine the upper limits of scaling on the photoBJT. By being able to independently optimize the performance of the gain region of the photoBJT, we can adopt the scale and doping levels used for the most ultimate performance BJTs published. One publication specifically tries to examine the upper limits of performance for SiGeC heterojunction BJT (HBT) technology, by simulating extremely scaled lateral dimensions with very aggressively scaled vertical junctions with highly engineered doping and mole-fraction grading of the Si and Ge components[2, 3]. Their results show rather impressive f_T values well over 1 THz. Although not pure germanium and missing the necessary dimensions for a reasonable absorption area, it is imaginable that modifications could be made to the structure to turn it into a photoBJT without dramatically degrading its performance, giving

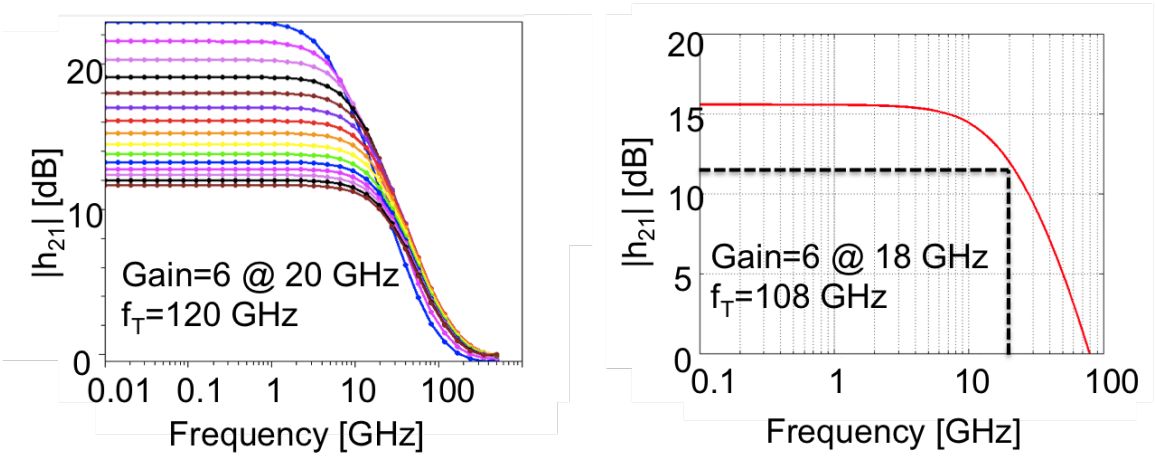

Figure 5.11: (a) Shows the simulated RF performance of the example photoBJT with separate gain and absorption regions. (b) Shows the gain versus frequency for the optimal bias point with excited with an optical signal.

some insight into a possible upper limit of a highly scaled photoBJT. As a point of reference, with a 1 THz gain-bandwidth product, such a photoBJT would be able to operate at a 100 GB/s bit rate with a gain of 10. The comparison between this theoretical upper limit, the photoBJT simulation which was previously described, and our preliminary data are shown in Fig. 5.12. A comparison with the state of the art APD is also shown.

5.7 Summary

In this chapter, we have presented several photoBJT designs, including a 32 GHz 3-terminal photoBJT, which was simulated and fabricated, although the initial fabricated devices showed that the emitter and collector were shorted, and additional calibration of the fabrication process is needed. However a 2-terminal floating base photoBJT was produced which was nearly identical in structure to the p-i-n photodiode in chapter 3, and which demonstrated a gan x bandwidth product of 14 GHz. Finally we considered how to appropriately scale a photoBJT, by separating the gain region from the absorption region, and an example 120 GHz photoBJT was simulated.

As we consider scaling the photoBJT in size to increase its performance, we must consider the limits for how far it can be scaled, not only in terms of the electrical performance, but the optical performance as well. Since germanium has a finite absorption length, eventually the device will be scaled so aggressively that very little light will be absorbed. To combat this issue, the next chapter focuses on creating a resonant cavity, coupled to a silicon waveguide, to enhance absorption in very small germanium devices.

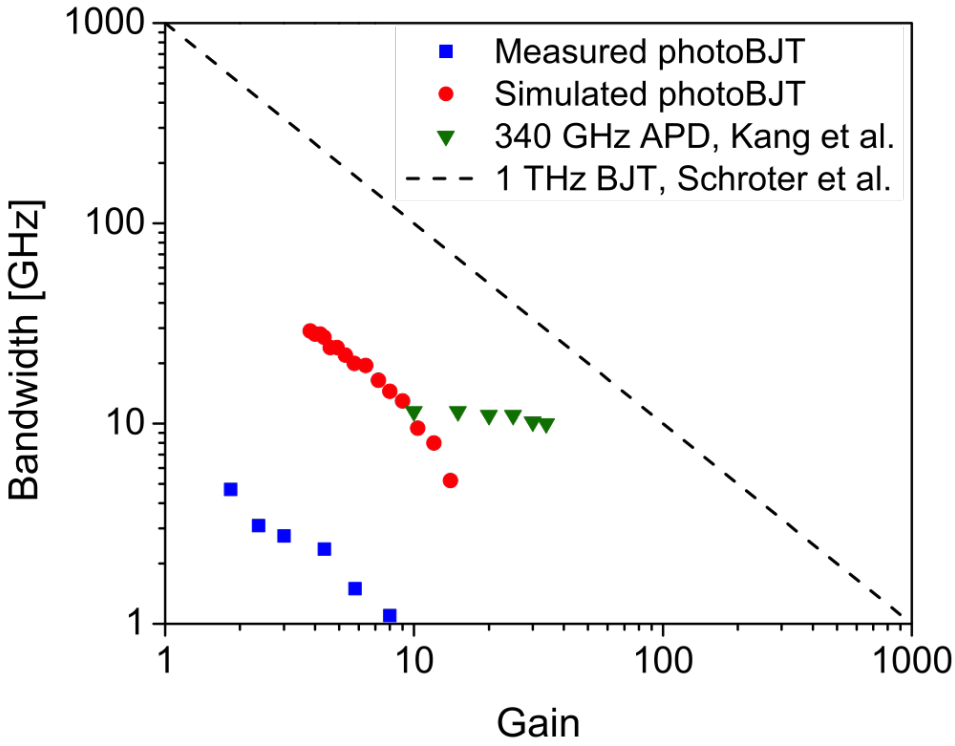

Figure 5.12: Gain-Bandwidth plot showing the measured results so far of our photoBJT, an example simulation of a future Ge photoBJT we could make, as well as a theoretical 1 THz photoBJT using the result from Schroter, et al[2, 3]. For comparison, the gain and bandwidth data are shown for the state of the art Ge/Si APD[4].

Chapter 6

Scaling to sub-fF Dimensions

While shrinking the physical dimensions of the photodiode will reduce its capacitance, doing so can severely reduce the quantum efficiency of the device by shrinking its absorptive area. This is compounded by the fact that the most likely material candidate for the photodiode is germanium due to its CMOS compatibility [88], however its absorption coefficient is relatively weak at 1550 nm [74]. The smallest capacitance photodiode currently reported is 1.2 fF with a quantum efficiency of nearly 67% [53], while a slightly higher capacitance diode, 2.4 fF, has also been developed but with much higher, 95%, quantum efficiency [89]. These are also some of the fastest reported germanium photodiodes since reducing RC delay increases overall speed [90].

Several methods have been used previously for enhancing the absorption in photodiodes such as creating resonant cavities [91] and metal antennas [92, 93]. However the resonant cavities previously described are still rather bulky, 130 fF capacitance [91], and rely on normal incident illumination, while modern silicon photonics utilizes silicon waveguides for light transport. The antenna coupled photodiodes, while having extremely low capacitance, 8 aF in one case, have extremely low quantum efficiency (<1%)[92]. This is partially due to issues related to pushing the limits of current fabrication, but also due to their fundamental reliance on plasmonic modes, which utilizes electrical energy being present in the metal of the antenna, which is extremely absorptive at optical frequencies. Metal-optic nanocavities can be an alternative, which have recently been proposed for efficient nanoscale lasers integrated onto silicon waveguides [94–98]. By relying on dielectric modes using the metal as a reflector, these cavities minimize metal losses, are extremely compact, and have been demonstrated to couple well to waveguide modes.

In this chapter, I will describe two aluminum-clad metal-optic nanocavities which strongly enhance light absorption in subwavelength germanium at 1500nm for sub-fF photodiodes. Detailed designs and simulations for both cavities are reported, along with a detailed analysis of their performance and design trade-offs using coupled mode theory. One design is for a metal semiconductor metal photodiode, with a capacitance of <200 aF and 39% external quantum efficiency. The second is a p-i-n photodiode with capacitance <100 aF and 51% external quantum efficiency. Both of these designs can easily be imagined to be extended

towards either a lateral npn phototransistor or vertical npn phototransistor respectively.

6.1 Tuning the optical cavity for optimal coupling

In order to optimize the structures for high quantum efficiencies, we simulated the devices using the finite difference time domain (FDTD) solver CST Microwave Studio. Both germanium and aluminum index of refraction and absorption coefficient values were taken from the Handbook of optical constants of solids [74, 99]. These values assume single crystal germanium, and aluminum evaporated or sputtered under high vacuum. Both simulated devices are air clad. To account for free carrier absorption associated with doping in the p-i-n device, both the polysilicon top contact, and bottom silicon waveguide are given an absorption coefficient of 300 cm^{-1}, which at 1500 nm, corresponds to a doping concentration of 5x10^{19} cm^{-3} [85].

In calculating the quantum efficiency of the device, we rely on two efficiency measures, η_{abs} and η_{coup}, which are the internal mode absorption efficiency and the external coupling efficiency, respectively. We define the quantum efficiency as the product of the two. We ignore in this analysis the internal electrical efficiency, which can reduce the total extracted photocurrent due to recombination from poor contacts, defect traps, and other imperfections in the electrical design. In the entirety of this chapter, all mentions of quantum efficiency assume that all absorbed photons produce electrons and holes that contribute to photocurrent. Thus the reported quantum efficiency describes the ratio of photons absorbed in the germanium divided by the number of photons incident on device from the silicon waveguide.

The internal mode absorption efficiency is calculated using coupled mode theory (CMT) [100] to calculate the absorption for a given mode using the relevant Q-factors. Starting from the CMT picture, let a be the mode amplitude in the resonator, and s^+ the incident wave amplitude. CMT allows us to describe the time varying mode energy[100] in Eqn. 6.1. The terms in the differential expression are respectively the phase procession of the mode, loss through re-radiation and absorption, and energy input through the waveguide. τ_{rad} and τ_{abs} are respectively the radiation rate of energy into and out of the cavity and absorption rate by the cavity. These can be written as a quality factor using the relation $1/\tau = \omega/Q$.

$$\frac{da}{dt} = j\omega_0 a - \frac{1}{2}\left(\frac{1}{\tau_{rad}} + \frac{1}{\tau_{abs}}\right)a + \sqrt{\frac{\eta_{coup}}{\tau_{rad}}}s^+ \tag{6.1}$$

Assuming a single frequency excitation at resonance, the equation can be solved for the steady-state solution, $\frac{da}{dt} = 0$, with the following expression of modal energy with

$$|a|^2 = \frac{\frac{\eta_{coup}}{\tau_{rad}}\left|s^+\right|^2}{\frac{1}{4}\left(\frac{1}{\tau_{rad}} + \frac{1}{\tau_{abs}}\right)^2} \tag{6.2}$$

Note that by definition $P_{inc} = \left|s^+\right|^2$, and so the absorbed power can be written as

$$P_{abs} = \frac{|a|^2}{\tau_{abs}} = \frac{4\frac{1}{\tau_{rad}}\frac{1}{\tau_{abs}}}{\left(\frac{1}{\tau_{rad}} + \frac{1}{\tau_{abs}}\right)^2}\eta_{coup}P_{inc} = \frac{4Q^2}{Q_{rad}Q_{abs}}\eta_{coup}P_{inc} \tag{6.3}$$

From inspection it can be seen that in order to maximize absorption by the cavity, $Q_{rad} = Q_{abs}$ since $1/Q = 1/Q_{rad} + 1/Q_{abs}$. This matching of radiation and absorption quality factors can be thought of as analogous to matching impedances in a circuit for maximal power transfer. Now we can write the expression for absorption efficiency in terms of the power absorbed by germanium over the total power absorbed by the structure, such as by the germanium and the metal.

$$P_{Ge} = \frac{|a|^2}{\tau_{Ge}} = \frac{4\frac{1}{\tau_{rad}}\frac{1}{\tau_{Ge}}}{\left(\frac{1}{\tau_{rad}} + \frac{1}{\tau_{abs}}\right)^2}\eta_{coup}P_{inc} = \frac{4Q^2}{Q_{rad}Q_{Ge}}\eta_{coup}P_{inc} \tag{6.4}$$

In the expression, $Q^{-1} = Q_{rad}^{-1} + Q_{metal}^{-1} + Q_{Ge}^{-1}$, and $Q_{abs}^{-1} = Q_{metal}^{-1} + Q_{Ge}^{-1}$. The internal mode absorption efficiency is then given by Eq. 6.5.

$$\eta_{abs} = \frac{4Q^2}{Q_{Ge}Q_{rad}} \tag{6.5}$$

To extract each Q-factor, the signal ring-down is calculated and absorptive loss is successively added. First the radiation Q-factor is calculated with a simulation where all materials are made artificially lossless. The metal Q is calculated from the same simulation which has been run again but with a realistic absorptive aluminum metal. The total Q of that simulation is the reciprocal sum of the radiation Q and the metal Q. Finally the germanium Q is calculated by running a third identical simulation with both metal loss and absorptive loss from the germanium. The total Q in that simulation is the reciprocal sum of the radiation Q, metal loss Q, and germanium absorption Q.

To obtain the external coupling efficiency, an additional simulation is run where instead of sending the excitation through the waveguide, an point-dipole emitter is placed within the cavity to excite the cavity mode, and the optical power leaving through the waveguide, P_{wg}, is calculated and compared with the amount of optical power leaving to total simulation space, P_{tot} such that $\eta_{coup} = P_{wg}/P_{tot}$. In this calculation, all materials are made artificially lossless. Thus the quantum efficiency, disregarding current extraction, defect traps, and other electrical imperfections, is the product, $\eta = \eta_{abs}\eta_{coup}$.

6.2 Lateral contacted cavity design

The proposed device, shown in Fig. 6.1(a), shows the metal clad cavity sitting on top of a terminated silicon waveguide. The primary feature seen is that the entire cavity is coated in aluminum, excepting a thin strip removed from the sidewalls and top, forming two electrically isolated metal walls which function as the contacts for the device. It sits on a 400 nm x 220

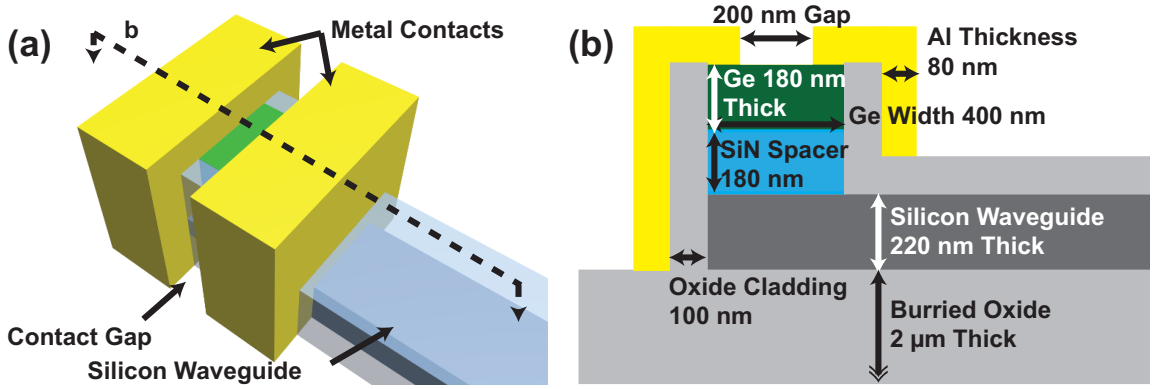

Figure 6.1: (a) 3-D drawing of the proposed device showing the metal-clad box sitting on the terminated input waveguide. Splitting down the center of the waveguide shows (b) the interior components of the cavity and device can be seen.

nm silicon waveguide on oxide. The waveguide couples evanescently through the bottom of the cavity. Figure 6.1(b) shows a cut-away running through the middle of the device along the waveguide. A small (400 nm x 400 nm x 180 nm) block of germanium is the absorber for the photodiode. The dimensions of the germanium are what primarily determine the cavity mode and the resonant wavelength. The split metal walls contact the top of the germanium directly for photocurrent extraction. The germanium can be undoped for a metal-semiconductor-metal photodiode, or laterally doped for a p-i-n photodiode. Surrounding the germanium on the sides is a thin (100 nm) layer of oxide, which isolates the mode in the germanium from the metal, reducing the metal absorption losses. Finally, the silicon nitride spacer controls the coupling between the waveguide and the cavity mode. By changing the thickness of this layer, critical coupling can be achieved, optimizing the absorption. This spacer can also be oxide, however silicon nitride was chosen for fabrication considerations. In both cases, it is possible to obtain relaxed and defect-free single-crystal germanium on this layer by rapid melt growth [73], which has been demonstrated to be relatively easy and reliable [52, 62]. To incorporate rapid melt growth, an island of silicon, electrically and optically isolated from the device, would be needed to act as the crystal seed. The removal of the excess germanium and seeding region would be a required part of the fabrication.

Figure 6.2(a) shows the cross-sectional mode profile of the proposed MSM photodiode structure seen in Fig. 6.1. The electric energy density is presented in dB scale to show the strong enhancement of the field intensity inside the germanium region with respect to the incident light in the waveguide. In the case shown the quantum efficiency is calculated to be 36% for a resonant wavelength of 1530 nm. In addition it can be seen that very little of the electric field extends into the metal.

The fundamental transverse-electric (TE) mode is operated in a given dimension, which originally has the doughnut electric-field profile, as shown in Fig. 6.2(b). Specifically the

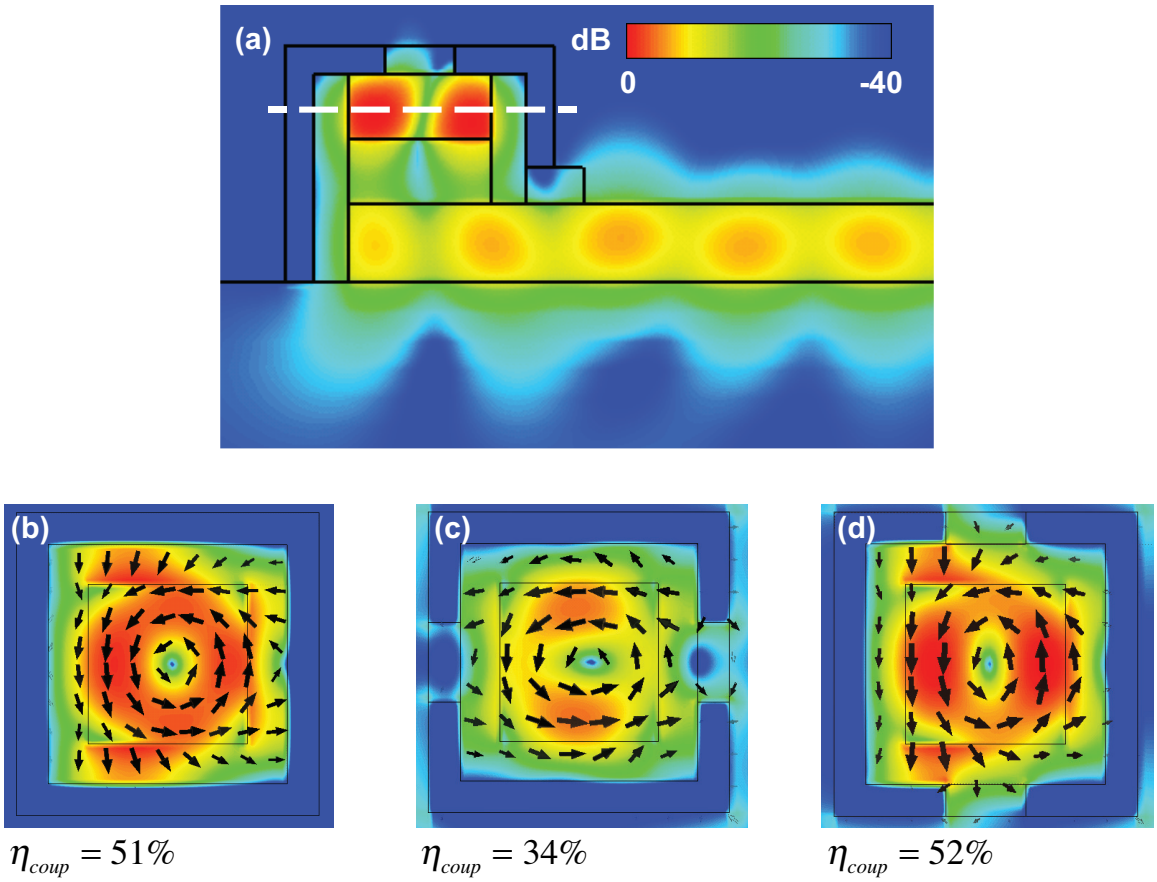

Figure 6.2: (a) Shows the electric energy density of the resonant mode. (b) Electric field profile with completely enclosed metal box. (c) Electric field profile with the metal gap parallel to the waveguide, and (d) perpendicular to the waveguide.

electric field is strongest in a doughnut ring around the interior of the germanium block. The electric field lines are parallel to the edge of the cavity. Because the germanium is electrically insulated from the silicon waveguide, all electrical contact must be made to the top or sides of the germanium. We examine the effect of separating the metal walls of the cavity to form top electrical contacts. In Fig. 6.2(b), the metal coating with no break is shown, and the mode is radially symmetric. In Fig. 6.2(c), the same mode is shown but with the break in the metal placed parallel to the waveguide, which causes the field to be enhanced parallel to the waveguide. This results in much poorer coupling between the waveguide and cavity mode, since the electric field lines of the cavity mode are mostly pointing perpendicular to the waveguide which leads to huge phase-mismatching. To fix this, the split in the metal is placed perpendicular to the waveguide, as shown in Fig. 6.1(a), and is the orientation used for the remainder of the paper. This orientation, shown in Fig. 6.2(d), enhances the electric field in the direction perpendicular to the waveguide, which is the same orientation

as that of the waveguide mode, which greatly enhances the coupling efficiency of the cavity by creating a better mode overlap. Two key parameters are examined in investigating both how to optimize the device as well as fabrication tolerances. Specifically these are the nitride spacer thickness, which controls the coupling strength to allow for critical coupling, and the spacing between contacts, which is a more critical issue for fabrication.

The primary consideration is in Q-matching the structure such that the radiation and absorption rates of the cavity are made equal. This is the very same concept used to create critical coupling in ring resonators to waveguides [101]. It is assumed the majority of the radiation is due to the coupling with the waveguide structure. Since the coupling distance is controlled directly by the thickness of the nitride spacer, altering this dimension in turn controls the radiation Q. Figure 6.3 shows the variation of the spacer layer thickness and its effect on radiation Q, absorption Q, and the overall quantum efficiency. It can be clearly seen that at the intersection of the radiation and absorption Q curves, the quantum efficiency reaches its peak of about 36%. In this case all other dimensions are as shown previously in Fig. 6.1(b). It should also be noted that as the spacer thickness becomes extremely thin, the cavity is over coupled to the waveguide, and the mode is not well defined, leading to less overall absorption and thus the higher absorption Q values. Once the cavity reaches critical coupling or under coupling, the absorption Q values remain relatively unchanged with respect to the spacer thickness. Overall this shows that the spacer thickness is fairly important to the proper coupling of the cavity mode, and that eliminating the spacer layer entirely would severely reduce the quantum efficiency.

It should be noted that the spacer material does not need to be nitride. For example, silicon oxide can be used instead, allowing the use of much more common GOI substrate in fabrication. However because oxide has a lower index of refraction than nitride, the optimal oxide thickness will be less (150 nm). A secondary effect is that the cavity resonance blue-shifts, requiring a change in the germanium dimensions to maintain a resonance at the desired wavelength.

One other consideration is the separation between the contacts. Shown in Fig. 6.4(a) is a variation of both the nitride spacer thickness and the contact separation. The total quantum efficiency is plotted, showing peaks due to critical coupling from the appropriate spacer thickness value. What can be seen is that while the quantum efficiency does increase slightly with a smaller gap and decrease slightly with a wider gap, overall the effect is that of a few percent in total efficiency, a total 4% change with a 150nm variation in the gap dimension. More importantly the critical coupling condition does not change appreciably. This is encouraging for two reasons. One is that it shows the gap can be made rather wide to allow for lower capacitance. For a 200 nm gap, the structure is calculated to have a total capacitance of 200 aF. If the gap is decreased to 100 nm, the capacitance would then double to 400 aF, while further reduction in the gap would only further increase the device capacitance. This increase in capacitance would negate the original purpose of the metal cavity design, that of creating a high quantum efficiency but extremely low capacitance photodiode. The other reason is that fabrication of such a device requires careful patterning of the sidewalls as well as that of the top. A slight variation in the gap, either on top, or on

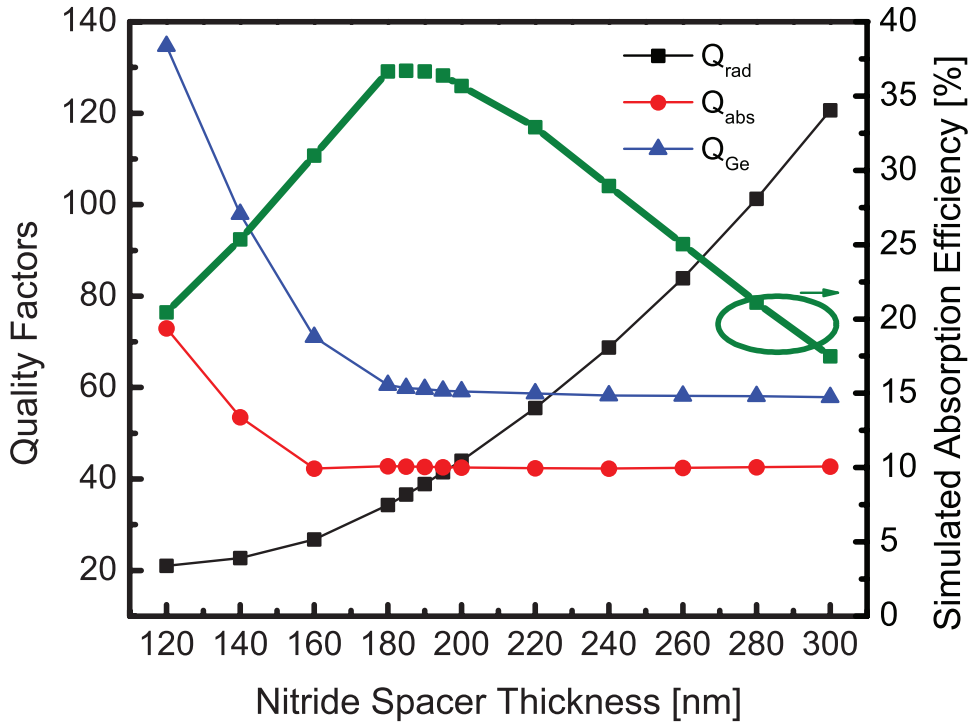

Figure 6.3: Simulated quality factors and total quantum efficiency for a variation in the nitride spacer thickness demonstrates the critical coupling effect in this structure.

the sides should not greatly affect the optical performance.

Another large point of variation in the structure is that of the germanium dimensions. Figure 6.4(b) shows the variation of length of the germanium block, while all other dimensions are kept constant. What is shown is that as expected the resonance red-shifts when the dimensions are lengthened, and blue-shifts when shrunk. The effect is rather linear, and a line of best fit can be applied to give an expression relating the resonant wavelength with the dimension. The fit shows that the cavity dimensions are roughly a half-wavelength of the resonant wavelength. This emphasizes that the resonance is the fundamental mode of a dielectric cavity and that the metal is simply acting as a reflector rather than creating a plasmonic mode.

6.3 Vertical contacted cavity design

The device drawn in Fig. 6.5 is very similar to that of the MSM structure with a few noticeable differences. Primarily, the device assumes vertical doping of the germanium, which requires contact through the bottom. By utilizing selective germanium growth on silicon, which has recently been developed to produce almost defect free germanium [53, 102], the bottom

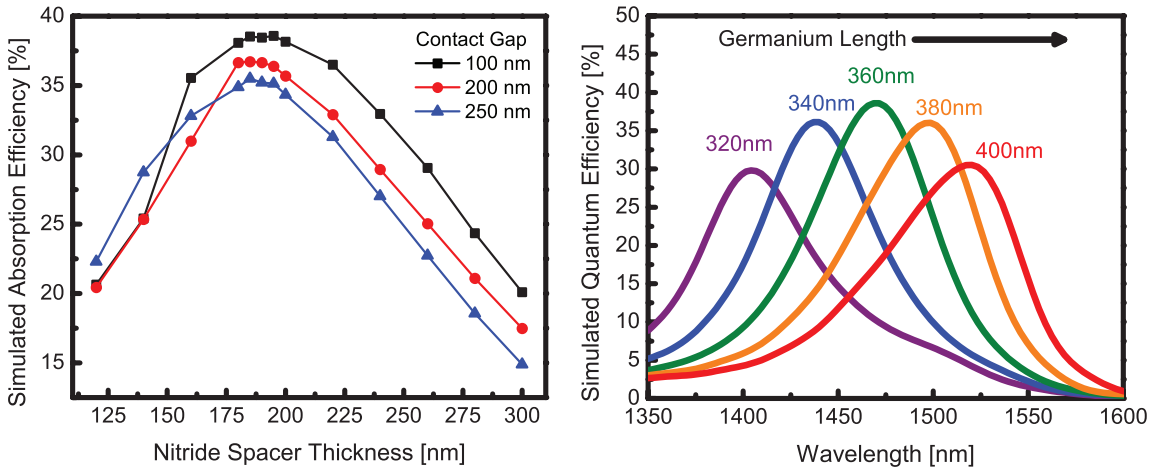

Figure 6.4: (a) Simulated quantum efficiency plotted against spacer thickness and metal gap. (b) Resonant wavelength versus the lateral germanium dimensions, showing how the wavelength can be tailored.

contact can in fact be made by the silicon waveguide. In order to make contact without disturbing the incoming waveguide mode, the waveguide is extended laterally from under the germanium region where aluminum is deposited on doped silicon, much like a truncated version of the waveguide integrated germanium photodiodes seen recently [53]. The metal box surrounding it is unbroken, and contacts the germanium indirectly through a poly-silicon layer, which serves to both isolate the mode from the metal and enhance the coupling to the cavity.

The dimensions of germanium are comparable to the previous design as seen from Fig. 6.5. One key difference is the increased germanium thickness. Because there is no longer an oxide or nitride layer separating the germanium from the waveguide, it must be much thicker than before to prevent overcoupling to the cavity. The oxide surrounding the Ge is 80 nm, and the metal is 150 nm thick, however it should be noted that with this design the metal can be as thick as desired, the only requirement being thicker than the skin depth. Finally the silicon contacts on the side of the device are 400 nm wide and about 750 nm long. It will be shown later that the length of these contacts is critical to the device performance because they act as additional waveguides for the cavity to couple to and thus much be chosen with care. Finally since this device is designed to be a p-i-n structure, both the germanium and the silicon must be doped, and since high doping are required for good contact, the free carrier absorption of silicon will begin to play a role in the device performance.

The p-i-n design has the same doughnut mode, whose electric field profile is identical to that of Fig. 6.2(a). However there are a few key differences which are displayed in Fig. 6.6. Primarily the coupling efficiency is much higher (>60%) compared to the MSM cavity for two reasons. In the MSM cavity, the light is radiated both through the gap between the metals, and through the base into the substrate. Thus much of light escapes before it can be

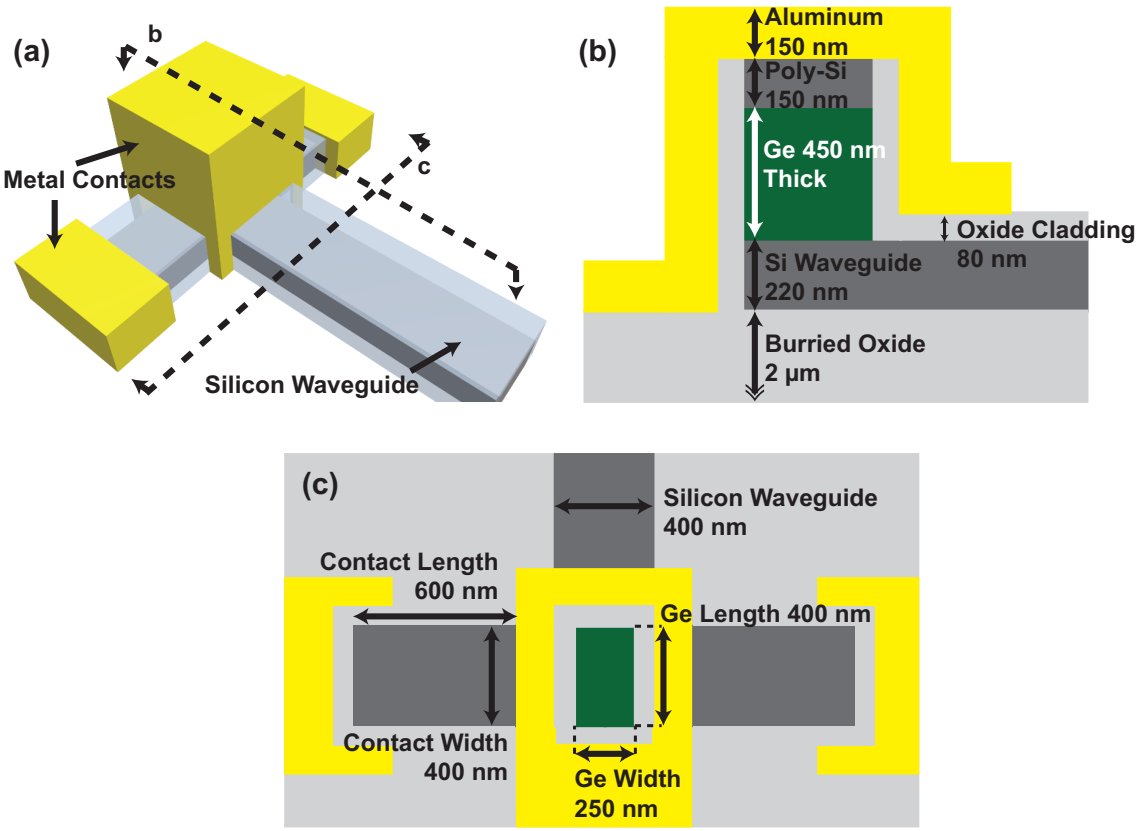

Figure 6.5: (a) 3-D drawing of the vertical p-i-n photodiode showing aluminum metal totally enclosing the active germanium. (b) A cut down the center of the input waveguide shows the internal cavity with top and bottom silicon contacts, while (c) a slice in the substrate plane reveals the lateral cavity dimensions as well as the side silicon contacts.

captured by the cavity mode. With the p-i-n design, there is no gap in the metal, eliminating one form of coupling loss. Because the p-i-n design has silicon waveguides on the side of the cavity for electrical contacts, a portion of the light that would enter the substrate actually couples into these waveguides, as in Fig. 6.6(b). If these contact waveguides are properly designed, this light when reflected back from them will couple very well to the cavity mode and is in a sense recycled. Further coupling improvements would likely require asymmetry in the oxide cladding on the sides of the cavity [94], something which is not realistic with present fabrication technology. Another important aspect seen in the p-i-n cavity is that the majority of the mode is present in the upper portion of the germanium. This is very important from a practical standpoint since with the current germanium growth technology, the first 50-100 nm of germanium grown on silicon is typically quite poor and thus would have many defect states present [53]. This defective layer of germanium could be doped so that it is not a part of the intrinsic absorbing layer.

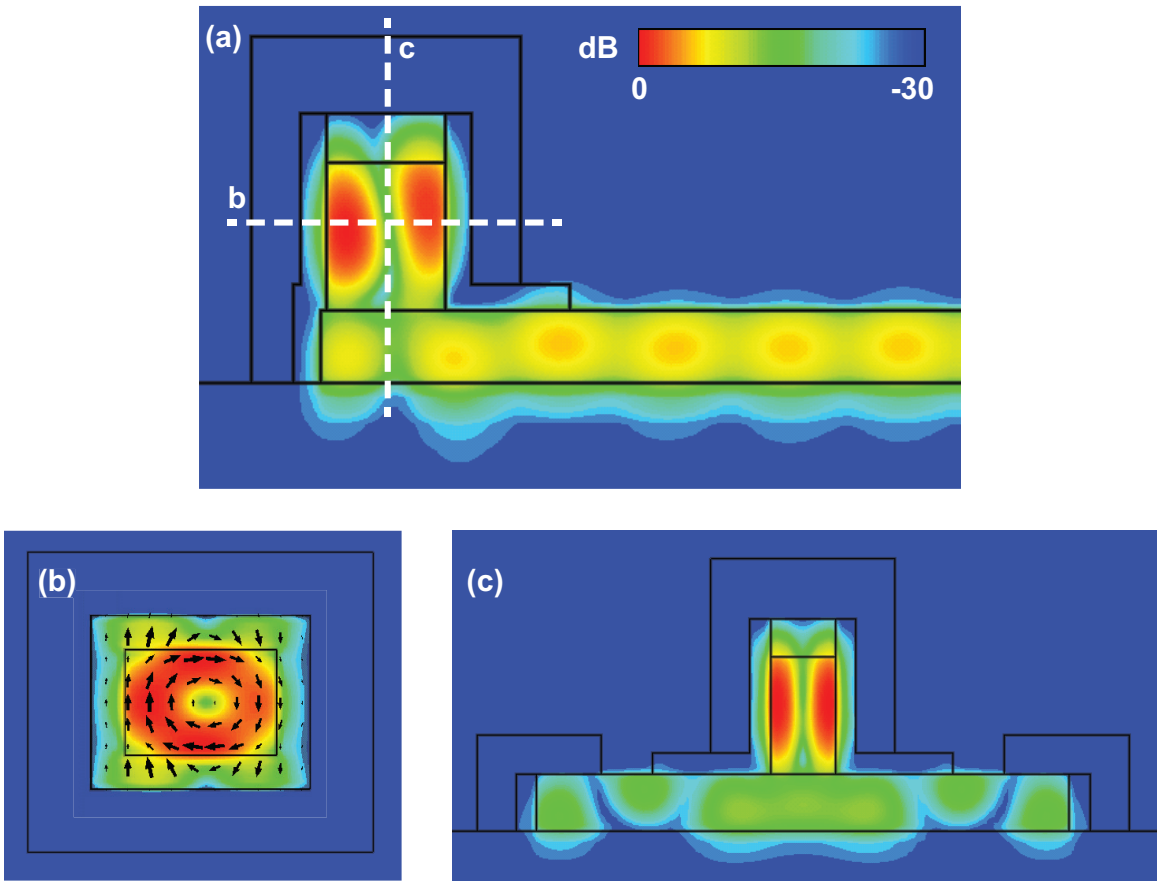

Figure 6.6: (a) Side-profile of the electric energy density in dB scale. (b) Electric energy density slice with electric field vectors showing similar doughnut mode. (c) Electric energy density of slice of the device showing the side silicon waveguide contacts.

To match the absorption Q for critical coupling there needs to be a way to adjust the radiation Q of the cavity. In the MSM design this was achieved with the nitride spacer thickness, however in the p-i-n design, there is no spacer between the germanium and silicon waveguide. In this case both the poly-silicon top contact and the total germanium thickness control the radiation Q such that increasing the germanium thickness will increase the radiation Q, shown in Fig. 6.7. What makes this more complicated than in the MSM design is that the total germanium dimensions are strongly tied to the resonance wavelength. Making the germanium thicker will red-shift the wavelength. Thus care must be taken to properly choose the length and width dimensions so that the proper radiation Q can be obtained at the desired wavelength. The fact that the length and width dimensions can alter the wavelength is shown in Fig. 6.8(b).

Additionally as in the MSM device, the cladding thickness strongly affects the coupling and the amount of mode overlap with the metal. A thinner cladding will provide better

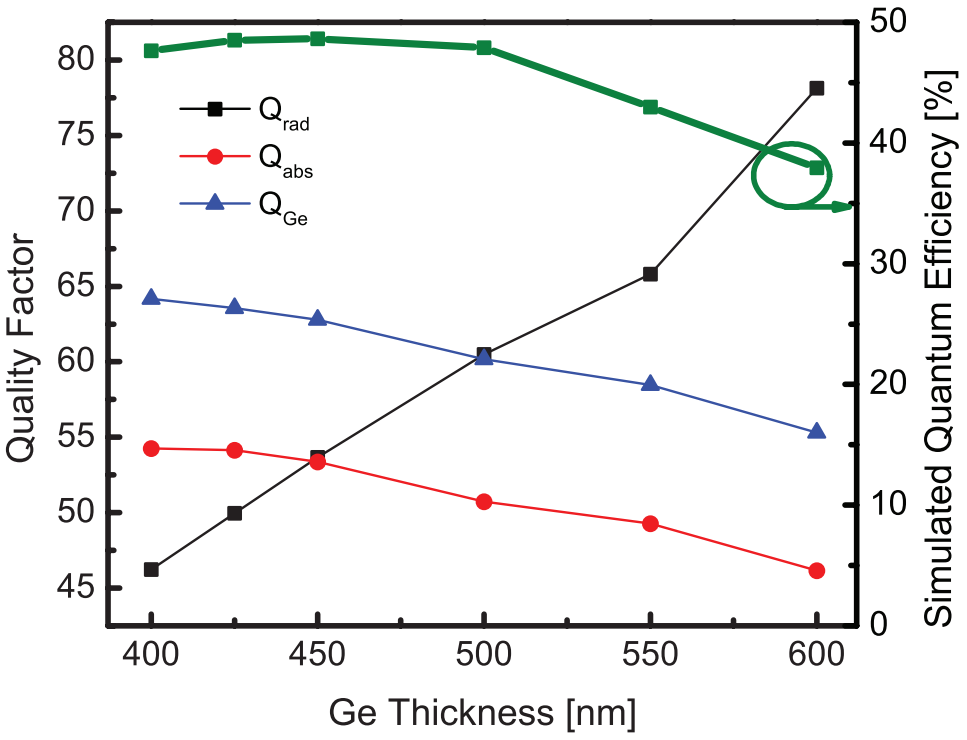

Figure 6.7: Matching the absorption and radiation Q by altering the germanium thickness.

coupling efficiency but at the cost of higher electric field penetration into the metal and thus higher metal loss. Thus a middle ground must be achieved between the two. For this cavity at 1500 nm wavelength that is about 80 nm.

As was previously mentioned, the length of the contact waveguides strongly affects the coupling efficiency in with the cavity. Specifically, the cavity also couples to these waveguides, and thus light from the input waveguide that does not immediately couple to the cavity mode with spread into these side waveguides. If their length is properly chosen, when this light reflects back from the end of them and reaches the cavity again, it will couple into the cavity. However if when the light makes its return trip from the end of the side waveguide it does not overlap with the mode then destructive interference will occur dramatically reducing the coupling efficiency. This is shown in Fig. 6.8(a). Taking this into account, it can be seen that by matching the Q factors, and choosing the proper contact length the p-i-n device can easily achieve over 50% quantum efficiency from simulations.

For a true p-i-n device, the polysilicon top contact, and underlying waveguide must both be heavily doped to allow good contact resistance. In the simulations it is assumed both silicon structures are doped to $5\text{x}10^{19}$ cm^{-3}, which gives a free carrier absorption of about 300 cm^{-1} at 1500 nm wavelength [85]. The excess loss caused by this doping was lumped into the Q_{metal} loss term, and indeed factors into the Q matching of the structure. However as can be seen from the electric energy density in Fig. 6, compared to the germanium itself,

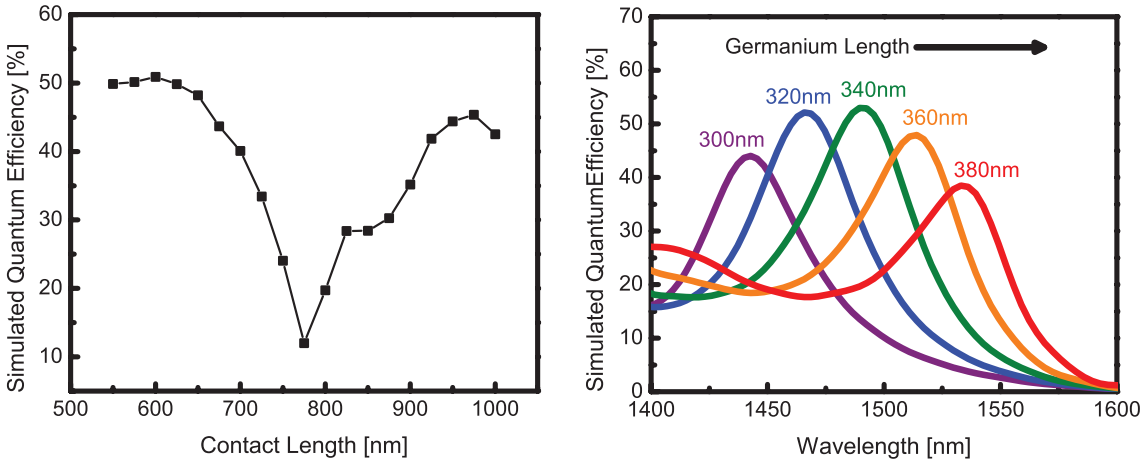

Figure 6.8: (a) Simulated quantum efficiency as a function of the silicon contact length dimensions as well as (b) how to tune the resonant wavelength by changing the germanium lateral dimensions.

very little energy is present in either part of the silicon. Combined with the fact that the germanium has an absorption coefficient of 2000 cm^{-1} at 1500nm, there is very little net effect from the silicon doping at these levels. Comparative simulations show a 2-3% increase in total absorption efficiency when the free carrier absorption is removed from the silicon in the structure. Additionally the reality of the film quality of the aluminum can affect absorption performance. In general, evaporation and sputtering can produce very smooth films, but aluminum is known to form a 2-5 nm interfacial oxide under even high vacuum conditions [99]. The effect of this thin Al_2O_3 layer between the aluminum and SiO_2 cladding the germanium is minimal. Aluminum oxide has a negligible absorption coefficient at 1500 nm, and the index of refraction (n=1.75) is comparable to SiO_2 (n=1.45). To design for such a layer, one would deposit 3-7 nm less SiO_2 around the device. However such considerations would need to be calibrated against specific aluminum deposition systems.

6.4 Summary

Two different but similar metal cavity designs were presented here. The first was a split metal cavity, which allows for a lateral doping and contact scheme, with very low capacitance ($<$ 200 aF), and high quantum efficiency (39 %). The second design was for a device where vertical doping and contacts are utilized, which would give lower capacitance ($<$ 100 aF) and higher quantum efficiency (51%). In either case, although the examples given were for a germanium photodiode, as was shown in the previous chapters, this could also be a germanium gate on a silicon body MOSFET or a germanium photoBJT depending on the doping of the germanium and the underlying structure of the silicon. In either case, this

accounts for silicon waveguide coupling, which will be needed to be incorporated in a silicon photonics platform. More importantly, while it may seem excessive to perform this level of engineering to extract slightly more absorption from a small block of germanium using metal, it is important to consider that when scaling the size of the photodetector, the metal is going to be present regardless for electrical contact purposes. It is thus important to consider then how the shape of the metal used for electrical connections can positively enhance the performance of the device rather than degrade its performance.

Chapter 7

Conclusions and Future Outlook

The previous chapters established the guidelines for implementing a highly energy efficient receiver by first examining a simply circuit model, pointing to the capacitance of the photodiode, and the length of the wire connection between it and the receiver amplifier as the bottleneck to ultimate sensitivity. By shrinking the photodiode and tightly coupling it to the first transistor stage, a highly sensitive, and thus low energy per bit receiver should be achievable. Silicon photonics was presented as a viable platform for this type of tight integration, not only because it lends itself towards tight integration of electronics and photonics, but also because entire photonic systems and their components have been demonstrated. To utilize this technology, and to achieve the tightest integration possible, this work has advocated for a germanium phototransistor design on silicon photonics.

To obtain high quality germanium on silicon photonics, the rapid melt growth technique has been presented, which allows for very thin, high crystal quality germanium to be grown on oxide. A germanium photodiode, wrapped around the silicon waveguide was shown to demonstrate the germanium quality and validate the fabrication process. The diode was found to have 6 nA of dark current, 0.95 A/W responsivity at 1550 nm, and 4 fF of total capacitance. The photodiode was found to be transit limited and with narrower designs operated at well over 40 GHz. All of these characteristics demonstrate a photodiode which meets or exceeds currently reported germanium photodiodes on silicon photonics.

The first phototransistor design envisioned was a germanium gated photoMOSFET, which has the advantages of being immediately implementable in standard MOS-based circuits. A 1 x 8 μm gate dimension photoMOSFET was fabricated and measured, and demonstrated up to 18 A/W responsivity with up to 2.5 GHz bandwidth. While the design had been previously proposed, there had been no demonstration of such performance yet. Although it was found the gain is nonlinear with input power, this may not be an issue if the circuit surrounding it is properly designed. Also because it is a typical MOSFET design other than the gate material, standard CMOS scaling rules can be applied, which was shown in simulation to dramatically improve responsivity and bandwidth.

The germanium bipolar phototransistor was seen as a device more likely to give both a high gain-bandwidth product and good linearity of gain. Several designs were presented, and

a floating base phototransistor was fabricated and demonstrated a 14 GHz gain-bandwidth product. The scalability of the germanium photoBJT was analyzed, and found that extremely high gain-bandwidth products in excess of 120 GHz can be achieved through having separate gain and absorption regions of the phototransistor. Scaling with these designs in mind could lead the photoBJT to overtake the avalanche photodiode in overall sensitivity and performance for future optical receivers.

Finally the issue of how physically small the germanium photodetector can be shrunk was addressed, since there is a finite absorption length, and eventually the device will become so small that the sensitivity gains from reduced capacitance will be eroded by the reduction in quantum efficiency of light absorption. Two metal cavity designs were described and simulated for both a lateral and vertical contact scheme, both of which would be coupled to a silicon waveguide. It was found in both cases, that a germanium photodiode could be fabricated in both cases with capacitance near 100 aF and quantum efficiency of over 50%. These cavities could equally be applied to either the germanium gate photoMOSFET or the germanium bipolar phototransistor as well.

In looking forward into the future, many directions present themselves. Because germanium devices will need to be thinner and more compact overall, the defect density near the germanium silicon interface for traditional epitaxial growth will become an increasing concern. Thus RMG may become a viable alternative, and as such it should be further explored and optimized for better uniformity, increased crystal quality, and ways to provide vertically contacted devices rather than just the laterally contact devices presented in this work.

The photoMOSFET may be further explored in the avenue of shrinking the gate dimensions and reducing channel length to increase speed and responsivity, although the speed may eventually be limited by the recombination lifetime within the germanium gate. One more interesting possibility would be to design an optical inverter circuit by creating both NMOS and PMOS complimentary photoMOSFETs, where the light would switch the output voltage digitally from off to on instead of relying on a nonlinear photocurrent gain.

The separated gain and absorption photoBJT is an extremely interesting device, and should be further explored through optimizing the design in simulation for faster gain-bandwidth products, and to actually realize the device through fabrication. Because the design lends itself to existing BiCMOS BJT designs, another interesting future route would be to create a germanium photoBJT sensitive to 1310 nm or even 850 nm light using an existing BiCMOS foundry process.

With either phototransistor, it would also be interesting to scale the germanium down to aF scale dimensions and utilize the metal nanocavity to resonantly enhance absorption for the ultimate low capacitance detector. This would then enable experimental testing of the sensitivity gains from low capacitance photodiodes tightly connected to a transistor.

Ultimately when any of these scaled phototransistor devices are successfully fabricated and tested, the most important test will be to determine their true sensitivity improvement. This would involve purchasing a commercial transimpedance amplifier (TIA) and wirebonding it to the fabricated device to create a more complete receiver circuit. Eye diagrams and bit error rate testing should be done, as well as some testing at elevated temperatures to

demonstrate performance even if the device were packaged into a computer case.

Overall this is just the start of very exciting work on new ultra-sensitive photoreceivers for ultra-low power optical interconnects. The explosion of future data centers and high performance computing will further drive the need for the development of these advanced technologies, and their development should have an enormous impact on the future of computing.

Bibliography

[1] A. Yariv and P. Yeh, ***Photonics: Optical Electronics in Modern Communications***, 6th ed. Oxford University Press, Jan. 2006.

[2] M. Schroter, J. Krause, N. Rinaldi, G. Wedel, B. Heinemann, P. Chevalier, and A. Chantre, "Physical and Electrical Performance Limits of High-Speed Si GeC HBTs—Part II: Lateral Scaling," *Electron Devices, IEEE Transactions on*, vol. 58, no. 11, pp. 3697–3706, 2011.

[3] M. Schroter, G. Wedel, B. Heinemann, C. Jungemann, J. Krause, P. Chevalier, and A. Chantre, "Physical and Electrical Performance Limits of High-Speed SiGeC HBTs—Part I: Vertical Scaling," *Electron Devices, IEEE Transactions on*, vol. 58, no. 11, pp. 3687–3696, 2011.

[4] Y. Kang, H.-D. Liu, M. Morse, M. J. Paniccia, M. Zadka, S. Litski, G. Sarid, A. Pauchard, Y.-H. Kuo, H.-W. Chen, W. S. Zaoui, J. E. Bowers, A. Beling, D. C. McIntosh, X. Zheng, and J. C. Campbell, "Monolithic germanium/silicon avalanche photodiodes with 340 GHz gain-bandwidth product," *Nature Photonics*, vol. 3, no. 1, pp. 59–63, 2009.

[5] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[6] C. Kachris and I. Tomkos, "A Survey on Optical Interconnects for Data Centers," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 1021–1036, Oct. 2012.

[7] D. A. B. Miller, "Device requirements for optical interconnects to silicon chips," *Proceedings of the IEEE*, vol. 97, no. 7, pp. 1166–1185, 2009.

[8] (2013) International Technology Roadmap for Semiconductors (ITRS). [Online]. Available: http://www.itrs.net/reports.html

[9] A. Krishnamoorthy, H. Schwetman, X. Zheng, and R. Ho, "Energy-Efficient Photonics in Future High-Connectivity Computing Systems," *Journal Of Lightwave Technology*, vol. 33, no. 4, pp. 889–900, Feb. 2015.

[10] A. K. Okyay, D. Kuzum, S. Latif, D. A. B. Miller, and K. C. Saraswat, "Silicon germanium CMOS optoelectronic switching device: Bringing light to latch," *Ieee Transactions On Electron Devices*, vol. 54, no. 12, pp. 3252–3259, Jan. 2007.

[11] S. Manipatruni, M. Lipson, and I. A. Young, "Device Scaling Considerations for Nanophotonic CMOS Global Interconnects," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 19, no. 2, 2013.

[12] S. B. Alexander, ***Optical Communication Receiver Design***. SPIE Publications, 1997.

[13] R. Soref, "The Past, Present, and Future of Silicon Photonics," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 12, no. 6, pp. 1678–1687, 2006.

[14] B. Jalali and S. Fathpour, "Silicon Photonics," *Lightwave Technology, Journal of*, vol. 24, no. 12, pp. 4600–4615, 2006.

[15] S. Assefa, H. Pan, S. Shank, W. M. J. Green, A. Rylyakov, C. Schow, M. Khater, S. Kamlapurkar, E. Kiewra, C. Reinholm, T. Topuria, P. Rice, C. Baks, and Y. Vlasov, "Monolithically integrated silicon nanophotonics receiver in 90nm CMOS technology node," in *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013*, 2013, pp. 1–3.

[16] H. Pan, S. Assefa, W. M. J. Green, D. M. Kuchta, C. L. Schow, A. V. Rylyakov, B. G. Lee, C. W. Baks, S. M. Shank, and Y. A. Vlasov, "High-speed receiver based on waveguide germanium photodetector wire-bonded to 90nm SOI CMOS amplifier ," *Optics Express*, vol. 20, pp. 18 145–18 155, Jul. 2012.

[17] X. Zheng, F. Liu, D. Patil, H. Thacker, Y. Luo, T. Pinguet, A. Mekis, J. Yao, G. Li, J. Shi, K. Raj, J. Lexau, E. Alon, R. Ho, J. E. Cunningham, and A. V. Krishnamoorthy, "A sub-picojoule-per-bit CMOS photonic receiver for densely integrated systems," *Opt. Express*, vol. 18, no. 1, pp. 204–211, 2010.

[18] Y. V. S. McNab, "Losses in single-mode silicon-on-insulator strip waveguides and bends," *Opt. Express*, vol. 12, no. 8, pp. 1622–1631, Apr. 2004.

[19] G. Li, J. Yao, H. Thacker, A. Mekis, X. Zheng, I. Shubin, Y. Luo, J.-h. Lee, K. Raj, J. E. Cunningham, and A. V. Krishnamoorthy, "Ultralow-loss, high-density SOI optical waveguide routing for macrochip interconnects," *Opt. Express*, vol. 20, no. 11, pp. 12 035–12 039, May 2012.

[20] A. Biberman, M. J. Shaw, E. Timurdogan, and J. B. Wright, "Ultralow-loss silicon ring resonators," *Optics Letters*, 2012.

[21] D. Taillaert, F. Van Laere, M. Ayre, W. Bogaerts, D. van Thourhout, P. Bienstman, and R. Baets, "Grating couplers for coupling between optical fibers and nanophotonic waveguides," *Japanese Journal of Applied Physics Part 1-Regular Papers Brief Communications & Review Papers*, vol. 45, pp. 6071–6077, 2006.

[22] A. Sure, T. Dillon, J. Murakowski, C. Lin, D. Pustai, and D. Prather, "Fabrication and characterization of three-dimensional silicon tapers," *Opt. Express*, vol. 11, no. 26, pp. 3555–3561, Dec. 2003.

[23] B. B. Bakir, A. V. de Gyves, R. Orobtchouk, P. Lyan, C. Porzier, A. Roman, and J. M. Fedeli, "Low-Loss (," *Ieee Photonics Technology Letters*, vol. 22, no. 11, pp. 739–741, 2010.

[24] B. Schmid, A. Petrov, and M. Eich, "Optimized grating coupler with fully etched slots," *Opt. Express*, vol. 17, no. 13, pp. 11 066–11 076, 2009.

[25] X. Chen, C. Li, C. K. Y. Fung, S. M. G. Lo, and H. K. Tsang, "Apodized Waveguide Grating Couplers for Efficient Coupling to Optical Fibers," *Ieee Photonics Technology Letters*, vol. 22, no. 15, pp. 1156–1158, 2010.

[26] Y. Tang, Z. Wang, L. Wosinski, U. Westergren, and S. He, "Highly efficient nonuniform grating coupler for silicon-on-insulator nanophotonic circuits," *Optics Letters*, vol. 35, no. 8, pp. 1290–1292, 2010.

[27] H.-Y. Chen and K.-C. Yang, "Design of a high-efficiency grating coupler based on a silicon nitride overlay for silicon-on-insulator waveguides," *Applied Optics*, vol. 49, no. 33, pp. 6455–6462, 2010.

[28] Y. Ding, C. Peucheret, H. Ou, and K. Yvind, "Fully etched apodized grating coupler on the SOI platform with -0.58 dB coupling efficiency," *Optics Letters*, 2014.

[29] T. Alder, A. Stohr, R. Heinzelmann, and D. Jager, "High-efficiency fiber-to-chip coupling using low-loss tapered single-mode fiber," *Ieee Photonics Technology Letters*, vol. 12, no. 8, pp. 1016–1018, 2000.

[30] Z. Xiao, X. Luo, P. H. Lim, P. Prabhathan, S. T. H. Silalahi, T.-Y. Liow, J. Zhang, and F. Luan, "Ultra-compact low loss polarization insensitive silicon waveguide splitter," *Opt. Express*, vol. 21, no. 14, pp. 16 331–6, 2013.

[31] P. D. Trinh, S. Yegnanarayanan, and B. Jalali, "Integrated optical directional couplers in silicon-on-insulator," *Electronics Letters*, vol. 31, no. 24, pp. 2097–2098, 1995.

[32] S. Han, T. J. Seok, N. Quack, B.-W. Yoo, and M. C. Wu, "Large-scale silicon photonic switches with movable directional couplers," *Optica*, vol. 2, no. 4, pp. 370–6, 2015.

[33] N. Sherwood-Droz, H. Wang, L. Chen, and B. G. Lee, "Optical 4x4 hitless silicon router for optical Networks-on-Chip (NoC)," *Opt . . .*, 2008.

[34] D. Feng, W. Qian, H. Liang, C.-C. Kung, J. Fong, B. J. Luff, and M. Asghari, "Fabrication Insensitive Echelle Grating in Silicon-on-Insulator Platform," *Photonics Technology Letters, IEEE*, vol. 23, no. 5, pp. 284–286, 2011.

[35] S. Cheung, T. Su, K. Okamoto, and S. J. B. Yoo, "Ultra-Compact Silicon Photonic 512 x 512 25 GHz Arrayed Waveguide Grating Router," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 20, no. 4, 2014.

[36] H. Nishi, T. Tsuchizawa, R. Kou, H. Shinojima, T. Yamada, H. Kimura, Y. Ishikawa, K. Wada, and K. Yamada, "Monolithic integration of a silica AWG and Ge photodiodes on Si photonic platform for one-chip WDM receiver," *Optics Express*, vol. 20, pp. 9312–9321, Apr. 2012.

[37] H. Rong, A. Liu, R. Jones, O. Cohen, and D. Hak, "An all-silicon Raman laser," *Nature*, vol. 433, no. 7023, pp. 292–294, 2005.

[38] O. B. B. Jalali, "Demonstration of a silicon Raman laser," *Opt. Express*, vol. 12, no. 21, pp. 5269–5273, Oct. 2004.

[39] R. E. Camacho-Aguilera, Y. Cai, and N. Patel, "An electrically pumped germanium laser," *Optics Letters*, 2012.

[40] J. Liu, X. Sun, R. Camacho-Aguilera, L. C. Kimerling, and J. Michel, "Ge-on-Si laser operating at room temperature," *Optics Letters*, vol. 35, no. 5, pp. 679–681, Jan. 2010.

[41] K. P. Homewood and M. A. Lourenço, "Optoelectronics: The rise of the GeSn laser," *Nature Photonics*, vol. 9, no. 2, pp. 78–79, Feb. 2015.

[42] G. Roelkens, L. Liu, D. Liang, R. Jones, A. Fang, B. Koch, and J. Bowers, "III-V/silicon photonics for on-chip and inter-chip optical interconnects," *Laser & Photonics Reviews*, vol. 4, no. 6, pp. 751–779, 2010.

[43] D. Liang, G. Roelkens, R. Baets, and J. E. Bowers, "Hybrid Integrated Platforms for Silicon Photonics," *Materials*, vol. 3, no. 3, pp. 1782–1802, Mar. 2010.

[44] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. J. Thomson, "Silicon optical modulators," *nature photonics*, vol. 4, pp. 518–526, Jul. 2010.

[45] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, and D. Rubin, "A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor," *Nature*, 2004.

[46] T. Baba, S. Akiyama, M. Imai, N. Hirayama, H. Takahashi, Y. Noguchi, T. Horikawa, and T. Usuki, "50-Gb/s ring-resonator-based silicon modulator," *Opt. Express*, vol. 21, no. 10, pp. 11 869–8, 2013.

[47] S. Ren, Y. Rong, S. A. Claussen, R. K. Schaevitz, T. I. Kamins, J. S. Harris, and D. A. B. Miller, "Ge/SiGe Quantum Well Waveguide Modulator Monolithically Integrated With SOI Waveguides," *Photonics Technology Letters, IEEE*, vol. 24, no. 6, pp. 461–463, 2012.

[48] L. Vivien, J. Osmond, J.-M. Fedeli, D. Marris-Morini, P. Crozat, J.-F. Damlencourt, E. Cassan, Y. Lecunff, and S. Laval, "42 GHz p.i.n Germanium photodetector integrated in a silicon-on-insulator waveguide," *Opt. Express*, vol. 17, no. 8, pp. 6252–6257, 2009.

[49] L. Vivien, A. Polzer, D. Marris-Morini, J. Osmond, J.-M. Hartmann, P. Crozat, E. Cassan, C. Kopp, H. Zimmermann, and J.-M. Fedeli, "Zero-bias 40Gbit/s germanium waveguide photodetector on silicon," *Opt. Express*, vol. 20, no. 2, pp. 1096–1101, 2012.

[50] D. Feng, S. Liao, P. Dong, N.-N. Feng, H. Liang, D. Zheng, C.-C. Kung, J. Fong, R. Shafiiha, J. Cunningham, A. V. Krishnamoorthy, and M. Asghari, "High-speed Ge photodetector monolithically integrated with large cross-section silicon-on-insulator waveguide," *Applied Physics Letters*, vol. 95, no. 26, 2009.

[51] N.-N. Feng, P. Dong, D. Zheng, S. Liao, H. Liang, R. Shafiiha, D. Feng, G. Li, J. E. Cunningham, A. V. Krishnamoorthy, and M. Asghari, "Vertical p-i-n germanium photodetector with high external responsivity integrated with large core Si waveguides," *Opt. Express*, vol. 18, no. 1, pp. 96–101, 2010.

[52] S. Assefa, F. Xia, S. W. Bedell, Y. Zhang, T. Topuria, P. M. Rice, and Y. A. Vlasov, "CMOS-integrated high-speed MSM Germanium waveguide photodetector," *Opt. Express*, vol. 18, no. 5, pp. 4986–4999, Jan. 2010.

[53] C. T. DeRose, D. C. Trotter, W. A. Zortman, A. L. Starbuck, M. Fisher, M. R. Watts, and P. S. Davids, "Ultra compact 45 GHz CMOS compatible Germanium waveguide photodiode with low dark current," *Opt. Express*, vol. 19, no. 25, pp. 24 897–24 904, Nov. 2011.

[54] H. Zhai, Z. Zhang, L. Li, and S. Ma, "Low-voltage high-speed thin-film-Ge surface PIN photodetectors integrated on Si waveguide," *Optical And Quantum Electronics*, vol. 41, no. 14-15, pp. 957–961, Jan. 2009.

[55] G. Li, Y. Luo, X. Zheng, G. Masini, A. Mekis, S. Sahni, H. Thacker, J. Yao, I. Shubin, K. Raj, J. E. Cunningham, and A. V. Krishnamoorthy, "Improving CMOS-compatible Germanium photodetectors," *Optics Express*, vol. 20, no. 24, pp. 26 345–26 350, Nov. 2012.

[56] J. Wang, W. Y. Loh, K. T. Chua, H. Zang, Y. Z. Xiong, S. M. F. Tan, M. B. Yu, S. Lee, G. Q. Lo, and D. L. Kwong, "Low-Voltage High-Speed (18 GHz/1 V) Evanescent-Coupled Thin-Film-Ge Lateral PIN Photodetectors Integrated on Si Waveguide," *Photonics Technology Letters, IEEE*, vol. 20, no. 17, pp. 1485–1487, 2008.

[57] C.-K. Tseng, W.-T. Chen, K.-H. Chen, H.-D. Liu, Y. Kang, N. Na, and M.-C. M. Lee, "A self-assembled microbonded germanium/silicon heterojunction photodiode for 25Gb/s high-speed optical interconnects," *Scientific Reports*, vol. 3, Nov. 2013.

[58] M. Casalino, G. Coppola, M. Iodice, I. Rendina, and L. Sirleto, "Near-Infrared Sub-Bandgap All-Silicon Photodetectors: State of the Art and Perspectives," *Sensors*, vol. 10, no. 12, pp. 10 571–10 600, 2010.

[59] S. Su, B. Cheng, C. Xue, W. Wang, Q. Cao, H. Xue, W. Hu, G. Zhang, Y. Zuo, and Q. Wang, "GeSn p-i-n photodetector for all telecommunication bands detection," *Opt. Express*, vol. 19, no. 7, pp. 6400–6405, Mar. 2011.

[60] J. Werner, M. Oehme, M. Schmid, and M. Kaschel, "Germanium-tin pin photodetectors integrated on silicon grown by molecular beam epitaxy," *Applied physics ...*, vol. 98, no. 6, p. 061108, 2011.

[61] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. Hoboken: John Wiley & Sons, 2007.

[62] S. Assefa, F. Xia, and Y. A. Vlasov, "Reinventing Germanium avalanche photodetector for nanophotonic on-chip optical interconnects," *Nature*, vol. 464, no. 7285, pp. 80–84, 2010.

[63] L. Virot, P. Crozat, J.-M. Fedeli, J.-M. Hartmann, D. Marris-Morini, E. Cassan, F. Boeuf, and L. Vivien, "Germanium avalanche receiver for low power interconnects," *Nature Communications*, vol. 5, Sep. 2014.

[64] C. R. Crowell and S. M. Sze, "Temperature dependence of avalanche multiplication in semiconductors," *Applied Physics Letters*, 1966.

[65] S. Sahni, X. Luo, J. Liu, Y.-h. Xie, and E. Yablonovitch, "Junction field-effect-transistor-based germanium photodetector on silicon-on-insulator," *Optics Letters*, vol. 33, no. 10, pp. 1138–1140, Jan. 2008.

[66] J. Wang, H. Zang, M. B. Yu, Y. Z. Xiong, G. Q. Lo, D. L. Kwong, and S. Lee, "Enhanced Sensitivity of Small-Size (With 1-mu m Gate Length) Junction-Field-Effect-Transistor-Based Germanium Photodetector Using Two-Step Germanium Epitaxy by Ultrahigh Vacuum Chemical Vapor Deposition," *Ieee Electron Device Letters*, vol. 30, no. 10, pp. 1066–1068, 2009.

[67] J. Wang, M. Yu, G. Lo, D.-L. Kwong, and S. Lee, "Silicon Waveguide Integrated Germanium JFET Photodetector With Improved Speed Performance," *Ieee Photonics Technology Letters*, vol. 23, no. 12, pp. 765–767, 2011.

[68] K.-W. Ang, M.-B. Yu, G.-Q. Lo, and D.-L. Kwong, "Low-voltage and high-responsivity germanium bipolar phototransistor for optical detections in the near-infrared regime," *Ieee Electron Device Letters*, vol. 29, no. 10, pp. 1124–1127, 2008.

[69] D. Liang and J. BOWERS, "Photonic integration: Si or InP substrates?" *Electronics Letters*, vol. 45, no. 12, pp. 578–581, 2009.

[70] D. Liang and J. E. Bowers, "Recent progress in lasers on silicon," *Nature Photonics*, vol. 4, no. 8, pp. 511–517, Jan. 2010.

[71] T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, and M. Bohr, "A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors," in *IEEE IEDM*. IEEE, 2003, pp. 11.6.1–11.6.3.

[72] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, 2nd ed. Cambridge University Press, May 2009.

[73] Y. Liu, M. Deal, and J. Plummer, "Rapid melt growth of Germanium crystals with self-aligned microcrucibles on Si substrates," *Journal of the Electrochemical Society*, vol. 152, no. 8, pp. G688–G693, 2005.

[74] R. F. Potter, "Germanium (Ge)," in *Handbook of Optical Constants of Solids*, E. D. Palik, Ed. Academic Press, Mar. 1985, pp. 465–478.

[75] L. Colace and G. Assanto, "Germanium on Silicon for Near-Infrared Light Sensing," *Photonics Journal, IEEE*, vol. 1, no. 2, pp. 69–79, 2009.

[76] H. Ye and J. Yu, "Germanium epitaxy on silicon," *Science and Technology of Advanced Materials*, vol. 15, no. 2, pp. 024 601–10, Mar. 2014.

[77] J. H. Nam, T. Fuse, Y. Nishi, and K. C. Saraswat, "Germanium on Insulator (GOI) Structure Using Hetero-Epitaxial Lateral Overgrowth on Silicon," *ECS Transactions*, vol. 45, no. 4, pp. 203–208, Apr. 2012.

[78] C. J. Tracy, P. Fejes, N. D. Theodore, P. Maniar, E. Johnson, A. J. Lamm, A. M. Paler, I. J. Malik, and P. Ong, "Germanium-on-insulator substrates by wafer bonding," *Journal of electronic materials*, vol. 33, no. 8, pp. 886–892, 2004.

[79] Y. Liu, M. D. Deal, and J. D. Plummer, "High-quality single-crystal Ge on insulator by liquid-phase epitaxy on Si substrates," *Applied Physics Letters*, vol. 84, no. 14, pp. 2563–2565, 2004.

[80] S. Balakumar, M. M. Roy, B. Ramamurthy, C. H. Tung, G. Fei, S. Tripathy, C. Dongzhi, R. Kumar, N. Balasubramanian, and D. L. Kwong, "Fabrication Aspects of Germanium on Insulator from Sputtered Ge on Si-Substrates," *Electrochemical and Solid-State Letters*, vol. 9, no. 5, p. G158, 2006.

[81] A. Abbadie, J.-M. Hartmann, and F. Brunier, "A Review of Different and Promising Defect Etching Techniques: from Si to Ge," *ECS Transactions*, vol. 10, no. 1, pp. 3–19, 2007.

[82] L. Chen and M. Lipson, "Ultra-low capacitance and high speed Germanium photodetectors on Silicon," *Opt. Express*, vol. 17, no. 10, pp. 7901–7906, 2009.

[83] S. Matsumoto and T. Niimi, "Concentration Dependence of a Diffusion Coefficient at Phosphorus Diffusion in Germanium," *Journal of the Electrochemical Society*, vol. 125, no. 8, pp. 1307–1309, Aug. 1978.

[84] N. A. DiLello, D. K. Johnstone, and J. L. Hoyt, "Characterization of dark current in Ge-on-Si photodiodes," *Journal Of Applied Physics*, vol. 112, 2012.

[85] D. Schroder, R. Thomas, and J. Swartz, "Free carrier absorption in Silicon," *IEEE Journal of Solid-State Circuits*, vol. 13, no. 1, pp. 180–187, 1978.

[86] M. Takenaka, K. Morii, M. Sugiyama, Y. Nakano, and S. Takagi, "Dark current reduction of Ge photodetector by GeO2 surface passivation and gas-phase doping," *Opt. Express*, vol. 20, no. 8, pp. 8718–8725, 2012.

[87] W. Shockley, M. Sparks, and G. K. Teal, "pn Junction Transistors," *Physical Review*, vol. 83, no. 1, pp. 151–162, 1951.

[88] M. Rouviere, M. Halbwax, J. Cercus, E. Cassan, L. Vivien, D. Pascal, M. Heitzmann, J.-M. Hartmann, and S. Laval, "Integration of germanium waveguide photodetectors for intrachip optical interconnects," *Optical Engineering*, vol. 44, no. 7, 2005.

[89] L. Chen and M. Lipson, "Ultra-low capacitance and high speed germanium photodetectors on silicon," *Opt. Express*, vol. 17, no. 10, pp. 7901–7906, Jan. 2009.

[90] J. Wang and S. Lee, "Ge-photodetectors for Si-based optoelectronic integration," *Sensors*, vol. 11, no. 1, pp. 696–718, 2011.

[91] O. Dosunmu, D. Cannon, M. Emsley, B. Ghyselen, J. Liu, L. Kimerling, and M. Unlu, "Resonant cavity enhanced Ge photodetectors for 1550 nm operation on reflecting Si substrates," *Ieee Journal Of Selected Topics In Quantum Electronics*, vol. 10, no. 4, pp. 694–701, 2004.

[92] L. Tang, S. E. Kocabas, S. Latif, A. K. Okyay, D.-S. Ly-Gagnon, K. C. Saraswat, and D. A. B. Miller, "Nanometre-scale germanium photodetector enhanced by a near-infrared dipole antenna," *Nature Photonics*, vol. 2, no. 4, pp. 226–229, 2008.

[93] F.-F. Ren, K.-W. Ang, J. Ye, M. Yu, G.-Q. Lo, and D.-L. Kwong, "Split bull's eye shaped aluminum antenna for plasmon-enhanced nanometer scale Germanium photodetector," *Nano Letters*, vol. 11, no. 3, pp. 1289–1293, 2011.

[94] M.-K. Kim, A. M. Lakhani, and M. C. Wu, "Efficient waveguide-coupling of metal-clad nanolaser cavities," *Opt. Express*, vol. 19, no. 23, pp. 23 504–23 512, 2011.

[95] M. T. Hill, Y.-S. Oei, B. Smalbrugge, Y. Zhu, T. de Vries, P. J. van Veldhoven, F. W. M. van Otten, T. J. Eijkemans, J. P. Turkiewicz, H. de Waardt, E. J. Geluk, S.-H. Kwon, Y.-H. Lee, R. Notzel, and M. K. Smit, "Lasing in metallic-coated nanocavities," *Nature Photonics*, vol. 1, no. 10, pp. 589–594, 2007.

[96] K. Ding, M. T. Hill, Z. C. Liu, L. J. Yin, P. J. van Veldhoven, and C. Z. Ning, "Record performance of electrical injection subwavelength metallic-cavity semiconductor lasers at room temperature," *Opt. Express*, vol. 21, pp. 4728–4733, Feb. 2013.

[97] M. T. Hill, M. Marell, E. S. P. Leong, B. Smalbrugge, Y. Zhu, M. Sun, P. J. van Veldhoven, E. J. Geluk, F. Karouta, Y.-S. Oei, R. Notzel, C.-Z. Ning, and M. K. Smit, "Lasing in metal-insulator-metal sub-wavelength plasmonic waveguides," *Opt. Express*, vol. 17, no. 13, pp. 11 107–11 112, Jan. 2009.

[98] Q. Ding, A. Mizrahi, Y. Fainman, and V. Lomakin, "Dielectric shielded nanoscale patch laser resonators," *Optics Letters*, vol. 36, no. 10, pp. 1812–1814, 2011.

[99] D. Y. Smith, E. Shiles, and M. Inokuti, "The Optical Properties of Metallic Aluminum," in *Handbook of Optical Contstants of Solids*, E. D. Palik, Ed. Academic Press, Aug. 1985, pp. 369–406.

[100] H. A. Haus, *Waves and Fields in Optoelectronics*. Prentice Hall, Aug. 1983.

[101] M. Soltani, S. Yegnanarayanan, Q. Li, and A. Adibi, "Systematic engineering of waveguide-resonator coupling for silicon microring/microdisk/racetrack resonators: theory and experiment," *Ieee Journal of Quantum Electronics*, vol. 46, no. 8, pp. 1158–1169, 2010.

[102] H.-Y. Yu, S. Ren, W. S. Jung, A. K. Okyay, D. A. B. Miller, and K. C. Saraswat, "High-efficiency p-i-n photodetectors on selective-area-grown Ge for monolithic integration," *IEEE Electr Device Lett*, vol. 30, no. 11, pp. 1161–1163, 2009.

Appendix A

Calibration and Setup of Optoelectronic S_{21} Measurement

All of the reported high speed measurements were performed with a network analyzer. Because the measurement is a relative response measurement, the calibration and setup of the measurement are very important to obtaining accurate results. This appendix will give specific details for the setup, calibration, and correction of the obtained S_{21} spectra for the p-i-n photodiodes, but can be extended to any frequency measurement described in this work, or for future measurements of photodetectors.

A.1 Setup and Calibration

Figure A.1 shows an overview schematic of the network analyzer setup. The basic setup involves the network analyzer (PNA), an external laser and modulator, and an RF probe for contacting the photodiode device under test (DUT). The external fiber laser is first connected to a polarization controller which is then connected into the electro-optic modulator. The modulator is polarization sensitive, and thus the polarization must be rotated to align it with the desired axis of the modulator. In this case we used an agilent tunable fiber laser, and an EO space 40 GHz $LiNbO_3$ intensity modulator. The output of the modulator is then connected to an optional erbium doped fiber amplifier (EDFA) which is used to boost the optical signal if needed. The output of the EDFA then goes through another polarization controller to align the polarization to that of the grating coupler on the silicon photonics chip.

Port 1 of the PNA is connected to the RF input of the EO modulator, which applies the small signal input to modulate the optical output of the modulator. A DC bias is applied separately to the modulator, and the output power of the modulator is monitored with the EDFA as the DC voltage is adjusted, so that it is half of the maximum output power. Because there will be some drift associated with the modulator, it is important to monitor the average output power level of the modulator and adjust the DC bias as necessary to keep

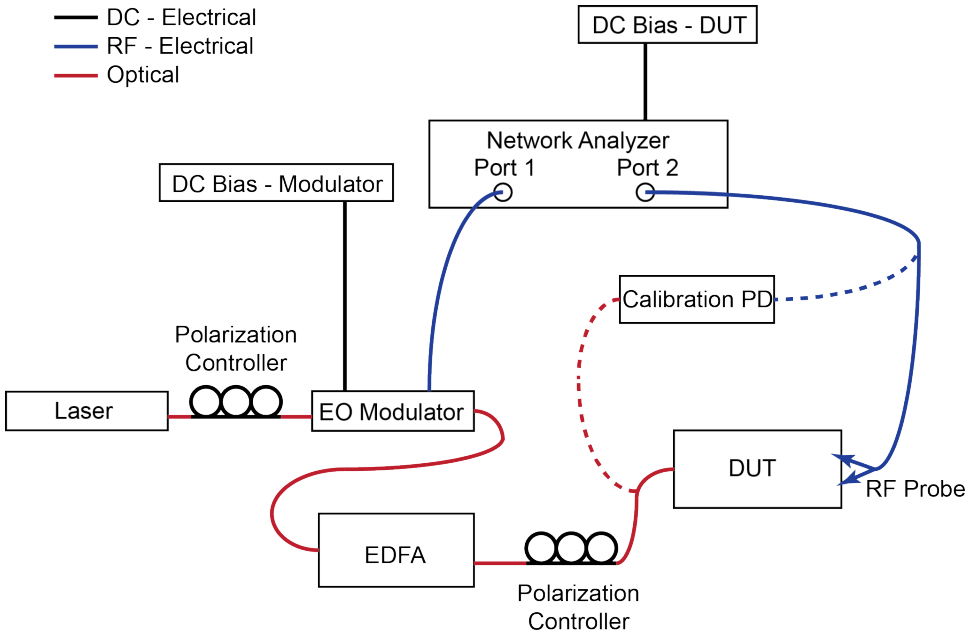

Figure A.1: Block diagram of the measurement setup for the S_{21} measurement.

the power close to half the maximum. This is the 'crossing point' of the modulator, and it is where its output is most linear with respect to applied voltage.

The DUT photodiode is probed with an RF probe, which has a coaxial connection that connects directly to port 2 of the PNA. In order to provide DC bias to the photodiode as well, a bias T is needed, but in this case, the PNA has a built-in bias T, and so the DC bias from a voltage supply is connected to the back of the PNA, where the DC connection for port 2 is located.

It is very important when making connections with coaxial connectors to use the torque wrench to tighten all connections to the appropriate tightness. This not only prevents damaging the connectors from over-tightening them, but more importantly it prevents having the connections be too loose. Even when connects are deemed 'finger tight', there can be a small gap remaining between the connectors, which creates an impedance mismatch and can result in reflections and ripples in the measured spectra.

Next the system must be calibrated so that we are only measuring the relative response of the DUT and not that of the modulator or RF cables. To do this we use a commercial photodiode, in this case a U2T 70 GHz photodiode, which accepts a single mode fiber on one end and a coaxial connection on the other, with separate DC bias contacts. The fiber at the output of the EDFA is connected to the diode, and the coaxial cable connected to the RF probe is connected to the diode as well. The DC bias is applied to the specifications of the diode. Then a 'thru' calibration is performed, which measures the response of the system and makes it flat across frequencies. The fiber and RF cable are then reconnected to the DUT.

When performing the calibration and measurement, at times the signal may not be very strong, and so to reduce noise in the measurement, typically several traces are taken and averaged together. Additionally, if there is a lot of random noise present, a small amount of smoothing may also be applied to the resulting spectra.

A.2 Extracting the 3dB Frequency

One we have obtained an S_{21} spectrum, we must add in the frequency responses which are not a part of the measured photodiode device, and which were not accounted for by the calibration. Namely the response of the RF probe was not accounted for with this type of calibration, so the probe response, which is generally documented when the probe is purchased, needs to be subtracted from the measured S_{21}. Additionally when calibrating the system, we inadvertently removed the frequency response of the calibration photodiode as well. In order to correct for this, we must add back the response of the commercial PD, which can be obtained from the data sheet that comes with the PD when it is purchased.

$$S_{21,DUTPhotodiode} = S_{21,measured} - S_{21,RFProbe} + S_{21,CalibrationPD} \tag{A.1}$$

To extract the 3dB frequency, because the spectrum is generally noisy or there may be ripples, a fit should be applied to the data. While given a purely RC response, it is well known the analytical function that describes frequency response, in a real device there may be multiple poles, and the response can be more complicated. To deal with this, it is much simple to apply a polynomial fit to the spectrum and extract the 3dB point from the fitted curve. Typically a 3rd or 4th order polynomial does a good job of capturing the overall shape of the spectrum, both at high and low frequencies, while still cutting through any noise or ripples present in the data.

A.3 Impedance mismatch and spectral ripples

If there is some impedance mismatch between the 50 ohm probes and system, reflections can occur, which will present themselves as ripples in the resulting S_{21} spectrum. If ripples are seen, it is important to first check that all of the coaxial connections are properly tightened with the torque wrench, and if any are loose, to tighten and re-perform the thru calibration. If this has all been checked, then the ripples may be due to impedance mismatch with the DUT, which was seen with devices in this work. While the polynomial fit can accurately extract the 3dB frequency even in the presence of these ripples, they are not intrinsic to the device itself, and may not be desirable if publishing the raw spectra. To correct for this, we make use of a post-processing technique employed in another paper[62]

What can be seen in Fig. A.2 are the spectra for the photodiodes mentioned in Ch. 3. These phodiodes were impedance mismatched with the RF probes, and so the spectra had ripples in them. Because the photodiodes were nearly identical and their impedance was

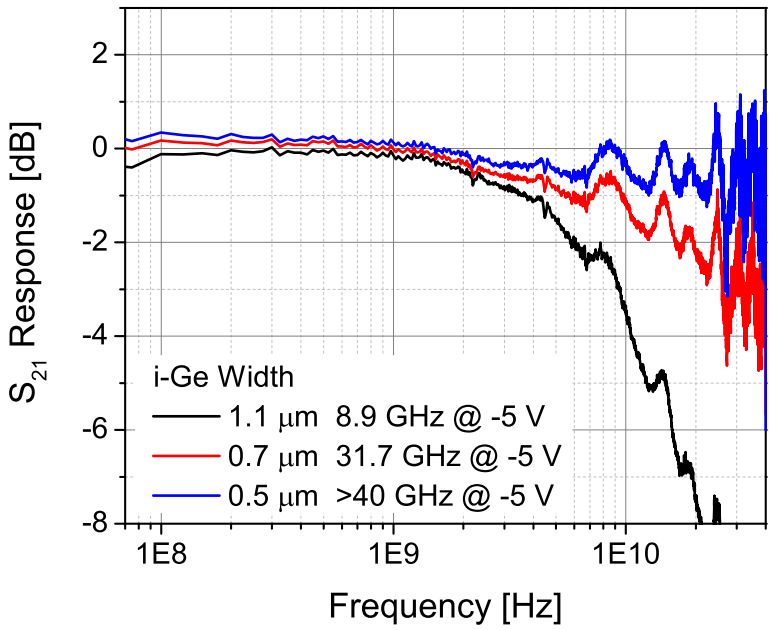

Figure A.2: Example of S_{21} spectra containing ripples due to impedance mismatch of the device.

nearly identical, the same ripples are seen in each curve for each device. To alleviate this, a polynomial fit is applied to one of the spectra. Which spectrum is fitted is not importance. The residuals are then extracted from the fitted curve, and the residuals are then heavily smoothed to remove any of the noise from the measurement itself, and all that is left should be the spectrum of the ripple itself. This can be seen in Fig. A.3.

The residual is then subtracted from each of the measured spectra, and the result are the spectra with the ripples removed, which can be seen in Fig. A.4. The extracted 3dB frequency is not affected.

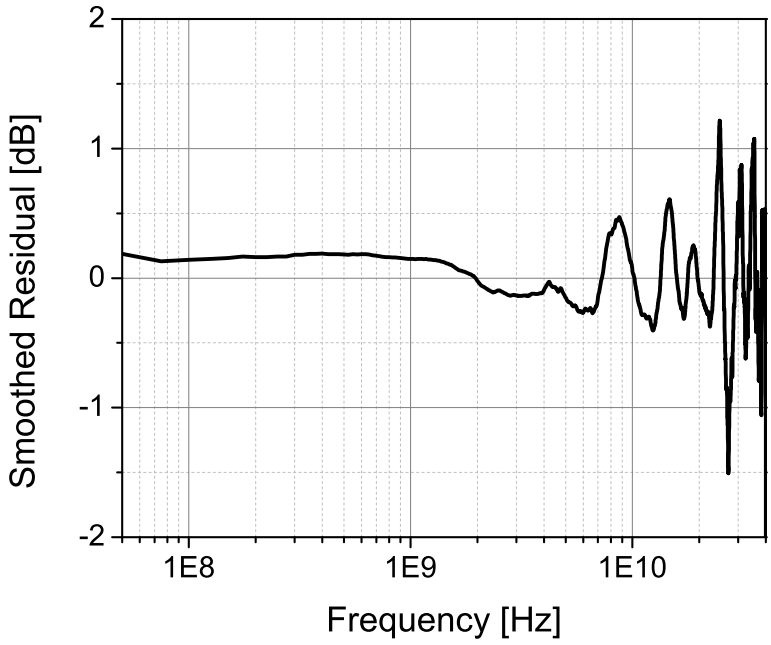

Figure A.3: Example of ripple from impedance mismatch which has been extracted from measured S_{21} spectra.

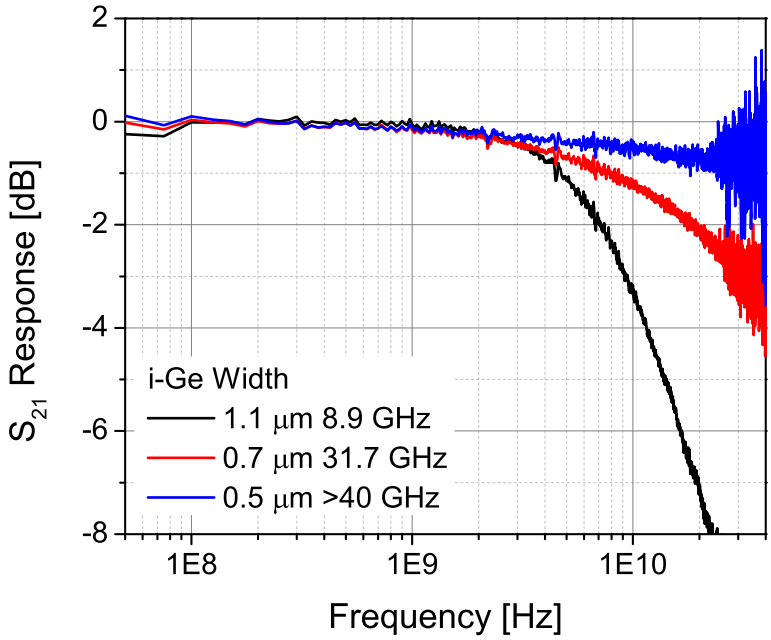

Figure A.4: Example of S_{21} spectra which have had the ripples from impedance mismatch removed.

Appendix B

PhotoBJT Runcard

Step #	Tool	Step	Recipe/ Details	Params	Comments
0	lam8	PM Etch	120 nm etch	OE-2 sec, ME-30s	Ran one wafer through ASML and recognized PM marks
1.1	svgcoat	Resist coat	"1-1-1"		
1.2	asml	FE layer		8 mJ	ran matrix, and examined under SEM. 8 mJ gives center waveguides 510 nm width in PR, get slightly wider towards edge of wafer.
1.3	svgdev	develop	"1-1-9"		
1.4	uvbake	hard bake	U		
1.5	lam8	full etch		ME - 60s, OE-10s	
1.6	matrix	resist strip			
2.1	svgcoat	Resist coat	"1-1-1"		
2.2	asml	PTE layer		15 mJ	ran matrix, and examined under SEM. 15 mJ gives ridge waveguides 495 nm width in PR, 50/50 duty cycle GC.
2.3	svgdev	develop	"1-1-9"		
2.4	uvbake	hard bake	U		

Step #	Tool	Step	Recipe/ Details	Params	Comments
2.5	lam8	partial etch		ME - 12s	Measured etch rate with dummy silicon, measuring depth with olympus. 60 nm target depth.
2.6	matrix	resist strip			
2.7	msink6	pirhana		10 min	
2.8	msink6	HF dip		10 s	25:1 HF
2.9	tystar2	oxidize	2dryoxa	1000C @ 14min, 5 min anneal	measured on nanospec: 20nm
3.1	svgcoat	resist coat	"1-2-1"		
3.2	asml	GSE layer		20 mJ	
3.3	svgdev	develop	"1-1-9"		
4.1	msink7	oxide etch	10:1 BHF	1.5 min	dummy wafer clear after 40s, let soak for another 40s to make sure all oxide gone.
4.2	msink1	resist strip	prs-3000	30 min	
5.1	msink8	pirhana		10 min	
5.2	msink6	pirhana		10 min	
5.3	msink6	HF dip	25:1 HF	10 s	
5.4	tystar20	Ge dep	sigenucf.20	seed: 300 mTorr, 100 sccm of Si2H6, 350C for 7.5 min, Dep: 600 mTorr, 88 sccm of GeH4, 350C for 60 min	
5.5	tystar11	LTO	11sulton	15s	Measured on nanospec: 18 nm
6.1	svgcoat6	resist coat	"1-2-1"		

Step #	Tool	Step	Recipe/ Details	Params	Comments
6.2	asml	GPD layer	10 mJ	ran matrix and looked in LEO. EbaseD measured 1.25 μm	
6.3	svgdev6	develop	"1-1-9"		
6.4	uvbake	hard bake	U		
6.5	lam8	Ge etch	rwgoing_Ge	10s oxide break-through, 60s main etch	oxide breaks about 10s into main etch, Ge endpoint starts 40s, and ends 50s, leaving about 10s of overetching. Measured oxide after etch, and overetching eats about 5nm of oxide.
6.6	matrix	resist strip			
7.1	msink7	hf dip	20 s		
7.2	msink7	h202	5 min @ 50C	strip Ge to rework wafer with new Ge. Need Ge coating waveguide during back-ground implant.	
7.3	msink6	pirhana			
7.4	tystar11	LTO	6.25E-2	measured 35 nm. Re-coat waveguide without consuming more Si	
8.1	svgcoat6	resist coat	"1-2-1"		

Step #	Tool	Step	Recipe/ Details	Params	Comments
8.2	asml	GSE layer		20 mJ	
8.3	svgdev6	develop	"1-1-9"		
8.4	msink7	oxide etch	10:1 BHF	1.5 min	
8.5	msink1	resist strip	prs-3000	30 min	
9.1	msink8	pirhana		10 min	
9.2	msink6	pirhana		10 min	
9.3	msink6	HF dip	25:1 HF	2s	
9.4	tystar20	Ge dep	sigenucf.20	seed: 300 mTorr, 100 sccm of Si2H6, 350C for 7.5 min, Dep: 600 mTorr, 88 sccm of GeH4, 350C for 60 min	
9.5	tystar11	LTO	15s	measured: 18 nm	
10.1	innovion implant				
11.1	svgcoat6	resist coat	"1-2-1"		
11.2	asml	GPD layer	10 mJ	ran matrix and looked in LEO. EbaseD measured 1.25 μm	
11.3	svgdev6	develop	"1-1-9"		
11.4	uvbake	hard bake	U		

Step #	Tool	Step	Recipe/ Details	Params	Comments
11.5	lam8	Ge etch	rwgoing-Ge	10s oxide break-through, 60s main etch	oxide breaks about 10s into main etch, Ge endpoint starts 40s, and ends 50s, leaving about 10s of overetching. Measured oxide after etch, and overetching eats about 5nm of oxide.
11.6	matrix	resist strip			
12.1	svgcoat6	coat	"1-2-1"		
12.2	asml	GCL	9 mJ		
12.3	svgdev6	develop	"1-1-9"	swab clean pm marks with acetone	
12.4	msink7	hf dip	10:1 BHF, 20 s	use dummy to verify	
12.5	msink7	h2o2	5 min @ 50C	actually removed GCL layer and used blanket resist. Coverage not good with GCL mask and dummy wafer had Ge structures removed	
12.6	matrix	resist strip			
12.7	msink7	10:1 BHF	20 s	remove all oxide from top of Ge	

Step #	Tool	Step	Recipe/ Details	Params	Comments
13.1	msink1	svc-14 clean			
13.2	tystar11	LTO	30 min, SULTON	measured 345 nm LTO	
13.3	rtp4	RMG	rwg1050.rcp	verified Ge is melted	
14.1	lam6	LTO etch	29 s, main etch	measured 65 nm after dry etch	
14.2	msink7	10:1 BHF	1 minute	used dummy to verify etch was finished.	
15.1	msink6	25:1 hf	10 s		
15.2	tystar11	LTO	5 s	Measured: 20 nm with nanospec.	
15.3	msink7	10:1 BHF	1s	Thinned LTO down to 4-5 nm, measured with nanospec.	
16.1	svgcoat6	resist	"1-2-1"		
16.2	asml	GEP	20 mJ	checked with LEO	
16.3	svgdev6	develop	"1-1-9"		
16.4	uvbake	hard bake	prog U		
16.5	innovion implant	check notes for implant energies and doses			
16.6	matrix	resist strip			
17.1	svgcoat6	resist	"1-2-1"		

Step #	Tool	Step	Recipe/ Details	Params	Comments
17.2	asml	GEN	20 mJ	checked with LEO	
17.3	svgdev6	develop	"1-1-9"		
17.4	uvbake	hard bake	prog U		
17.5	innovion implant	check notes for implant energies and doses			
17.6	matrix	resist strip			
18.1	msink1	svc-14	10 min	pre-furnace clean	
18.2	tystar11	11SULTON	23.5 min	measured with nanospec: 278-284 nm	
19.1	svgcoat6	resist	"1-1-1"		
19.2	asml	CON	21 mJ	checked with LEO	
19.3	svgdev6	develop	"1-1-9"		
19.4	uvbake	hard bake	prog U		
19.5	lam6	oxide etch	29 s main etch, 6 s overetch	calibrated main-etch rate with blanket oxide wafer, 10 nm/s	
19.6	matrix	resist strip			
20.1	mrc944	metal	rwgtial recipe	expected 50 nm Ti, 350 nm Al	
21.1	svgcoat6	resist	"1-2-1"		
21.2	asml	MET	11 mJ	checked with LEO	
21.3	svgdev6	develop	"1-1-9"		
21.4	uvbake	hard bake	prog U		

Step #	Tool	Step	Recipe/ Details	Params	Comments
21.5	lam7	metal etch	Main etch	45 s	Al etched 20s mark, and Ti etched 25s mark. 20s of overetching removes 100 nm of LTO but want to clear sidewalls of any remaining metal.