

Highly integrated Germanium Photo-detectors and III-V
Hybrid Lasers for Silicon Photonic Applications

Subal Sahni

Table of Contents

Table of Contents.....	2
Chapter 1: Introduction.....	5
1.1 Motivation.....	5
1.2 Photodetectors on Silicon.....	7
1.3 Lasers on Silicon	10
1.4 Organization of this dissertation	12
Chapter 2: Deposition and Characterization of Germanium Thin Films on Silicon: Theory.....	14
2.1 Issues with Ge growth on Silicon	14
2.2 Interfacial states and band alignment in Ge/Si	16
2.3 Photo-carrier transport across Ge/Si junction.....	18
2.4 Extraction of diffusion length from responsivity measurements	19
2.5 Epitaxial techniques for dislocation reduction.....	22
Chapter 3: Deposition and Characterization of Germanium Thin Films on Silicon: Experimental Results	25

3.1	<i>Ge deposition details</i>	25
3.2	<i>Direct characterization of the Ge films</i>	29
3.3	<i>Device fabrication and experimental setup</i>	33
3.4	<i>Diffusion length measurements: Results and discussion</i>	37
3.5	<i>Dependence of responsivity on wavelength</i>	43
3.6	<i>Conclusions: Optimized deposition conditions</i>	44
Chapter 4: Design and Demonstration of a Ge/Si Photo- Heterojunction FET		46
4.1	<i>Design Principle</i>	46
4.2	<i>Ge-on-SOI: Epitaxy and p-n junction characterization</i>	51
4.3	<i>Initial devices: Fabrication and results</i>	57
4.4	<i>Self-aligned detectors: Fabrication</i>	67
4.5	<i>Results: Pulsed response</i>	78
4.6	<i>Analysis of detector response</i>	83
4.7	<i>Conclusions: Why use Photo-HFETs?</i>	101
Chapter 5: Design of a Hybrid III-V/Silicon Evanescent Laser		108
5.1	<i>Basic design</i>	108
5.2	<i>Numerical Analysis</i>	119
5.3	<i>Towards experimental demonstration</i>	131

5.3	<i>Conclusions</i>	139
Chapter 6:	Conclusions	141
Appendix A:	Estimating the band-gap and refractive index of AlGaAsSb	145
Appendix B:	Analytic solutions for the modes of a disk resonator	148
Appendix C:	Estimating disk-waveguide coupling strength analytically	150
References	152

Chapter 1: Introduction

1.1 Motivation

As micro-electronic systems continue to scale down, the resultant electrical interconnect density would have to accommodate increased power dissipation, signal delay and cross-talk¹⁻³, while running at extremely high clock speeds. Current trends indicate that in less than a decade the power consumed by interconnects might become the limiting factor in determining the switching speeds in integrated circuits. To overcome these problems the integration of optical interconnects and systems alongside conventional Silicon (Si) based micro-electronic circuits would be a major step forward. Light based intra and inter-chip communication would diminish latency considerably² and reduce power consumption by eliminating capacitive interconnect loss³. Improved opto-electronic integration can also help increase the speed of transceiver circuits used in optical communication systems, thereby increasing the overall bandwidth.

Since CMOS (Complementary Metal-Oxide-Semiconductor) is the well-established technology utilized for the fabrication of electronic circuits, any such opto-electronic integration would require compatibility with CMOS. The challenge in photonics therefore, is to develop materials and processing techniques for photonic devices that can be seamlessly conjoined with existing CMOS based electronics. The resulting Si based photonics architecture would take advantage of the CMOS based

facilities available in the micro-electronics industry, thereby reducing fabrication costs and improving yield.

The utility of CMOS compatible fabrication would be enhanced if it avoided major changes in standard foundry processes because that would lead to easy adoption and widespread use of the proposed nano-photonic devices. For instance, a post-/back-end processing route, which refers to the fabrication of photonic devices on top of existing CMOS structures, might be preferable. However, there is a major constraint on the processing temperature: back-end processing is usually restricted to less than 450 °C to avoid degradation of the Al/Ti based contacts and interconnects. Making efficient photonic devices under this temperature constraint is a major technological barrier and requires extensive innovation in both fabrication techniques and device design.

With the goals mentioned above in mind, the research reported in this dissertation deals with the development of Si based photo-detectors and lasers that can be easily integrated onto standard Silicon chips. The majority of the work is on the design and fabrication of Germanium photo-detector structures that adhere to the temperature constraints described above, yet perform to specifications that are suitable for use in an actual receiver circuit.

To a lesser extent novel Silicon based laser designs are explored, predominantly on a theoretical level. The goal is to create an integrated device that can replace the discrete laser components that are bonded onto current opto-electronic chips.

1.2 Photodetectors on Silicon

The detection of light in a Si based setting is a very challenging problem. The large band-gap of Silicon (1.1 eV) makes it inactive in the wavelength range used in fiber-optic communication systems (1.3-1.55 μm , or photon energies ranging from 0.8 to 0.95 eV). In this wavelength window, certain III-V direct band-gap semiconductors have very high efficiency and speed of operation⁴. However, due to extremely large lattice mismatches and concerns over contamination issues, their integration with Si is extremely difficult. While the indirect band-gap of the elemental semiconductors precludes their use as efficient light emitters, one does not need very high efficiencies on the detector side, and therefore Germanium, with an indirect gap of 0.66 eV, can function well as a photo-detector. But even though Ge is more compatible with Si than the III-V semiconductors, good quality Ge growth on Si still demands high temperature processing (>700 °C)^{5,6,7}. The use of high temperature fabrication techniques causes difficulty in integrating the growth sequence with a practical back-end fabrication process, and can also lead to impurity segregation and diffusion, in addition to very significant inter-mixing between the Ge and Si.

The challenge, as mentioned earlier, is to make a photo-detector at low temperatures (< 450 °C), which gives adequate performance and can be integrated with Si efficiently. It must be mentioned that this temperature constraint isn't necessarily binding when it comes to the introduction of Germanium into CMOS: SiGe structures are already being incorporated into Silicon MOSFETs in order to introduce strain. Therefore some

foundries have inserted Ge epitaxy into the middle of their process flows and hence the temperature constraint isn't as stringent as the one on back-end processing. But in spite of the slightly higher thermal budget that might be available in an actual fab, it is extremely informative to look at the harder problem first, i.e., making the devices at a temperature that is back-end compatible. The findings can then be generalized and adapted to the device fabrication in an actual foundry. As will be discussed in subsequent chapters, certain techniques that do not adhere strictly to the 450 °C limit, were also briefly explored in this project.

1.2.1 Experimental Goals

It has been shown previously that electron-beam evaporation of polycrystalline Ge on Si substrates^{8,9} at 300 °C yields photo-detectors with moderate performance. In order to improve upon and extend the above work with low temperature Ge deposition, fundamental issues involving deposition rate, temperature and Si substrate surface preparation have to be addressed.

The first part of our work investigates these parameters using Ge thin-film deposition techniques including electron-beam evaporation (e-beam) and molecular beam epitaxy (MBE). The deposition experiments are carried out on both bulk Silicon and SOI. Various substrate cleaning techniques are adopted to try and improve the crystallinity of the epitaxial Ge. The quality of the Germanium is gauged by measuring the responsivity of simple p-n junction photo-detectors fabricated from the thin films. The bulk of the work on photo-detectors then involves utilizing the films deposited on SOI in actual devices that can be used in practical applications. SOI is the substrate of choice for

making the integrated devices because it lends itself extremely well to the waveguide based photonics that is utilized in any integrated opto-electronic chip. The goal in our research is to use device design to overcome the limitations imposed by the poor quality of the Ge films that is inherent to a low temperature growth process.

Keeping that in mind, we demonstrate a novel Ge/Si Photo-Heterojunction FET (Photo-HFET) that represents a completely new approach to looking at light detection in a Silicon platform. The design consists of a Si FET wherein the conventional gate is replaced by a Germanium island on the Silicon channel (Fig 1-1).

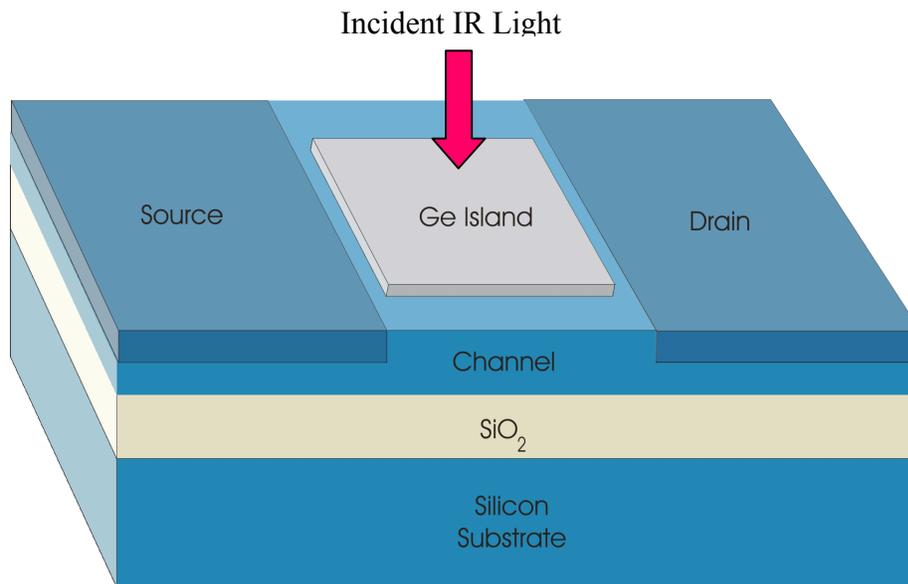


Figure 1-1: Geometry of a Ge/Si Photo-Heterojunction FET. The Germanium gate switches the conductance of the channel in response to incident infrared light.

When light is incident on the Ge, it modulates the conductance of the Si underneath which is measured as a change in channel current. The detector therefore behaves like an FET whose switching characteristics are governed by the light that is incident upon it. Chapters 4 describe the device in much greater detail.

1.3 Lasers on Silicon

The reduction in the cost of making highly integrated opto-electronic chips mentioned in the first section would be greatly enhanced if lasers could be monolithically integrated alongside the other components on the Si chip. The present practice is to in-couple the light from a discrete external source that may or may not be bonded onto the chip. The bonding process is clearly not a VLSI technology and can significantly increase the cost and turnaround time in the production of the photonic chip. The obvious solution therefore is to fabricate the laser on the Silicon die itself. But to achieve that, the problem of the indirect band-gap of Silicon must again be overcome. The lack of a direct gap makes the efficiency of radiative recombination in Si extremely low. Due to this, researchers have resorted to using Erbium doped Si-rich oxides¹¹⁻¹⁴, Silicon nanocrystals¹⁵⁻¹⁷ or surface-textured bulk Si¹⁸ to achieve light emission from Silicon. Even though these technologies are promising, they are not well suited to a highly integrated CMOS nano-photonic platform.

Another technique that has recently been developed takes advantage of the high Raman gain coefficient in Silicon to create a laser in the 1.55 μ m window^{19,20}. Even though excellent results have been reported in these experiments, the need for an optical pump beam means that these devices don't represent a complete solution to the problem of having a monolithically integrated light source on chip.

Keeping all these issues in perspective we propose and analyze a hybrid Silicon/III-V laser that is based on evanescent coupling. III-V alloys are the preeminent

materials in the photonics industry since they exhibit direct band-gaps that can be tuned to any wavelength by picking the appropriate composition. Unfortunately, as mentioned in the previous section, their integration onto Si is totally impractical. We therefore take a novel approach to utilizing III-V materials for light emission on Silicon: even though the III-V film does act as the gain medium, the cavity for the laser is made on Silicon. The coupling between the light in the Silicon and the gain medium is through the evanescent tail of the cavity mode (Fig 1-2).

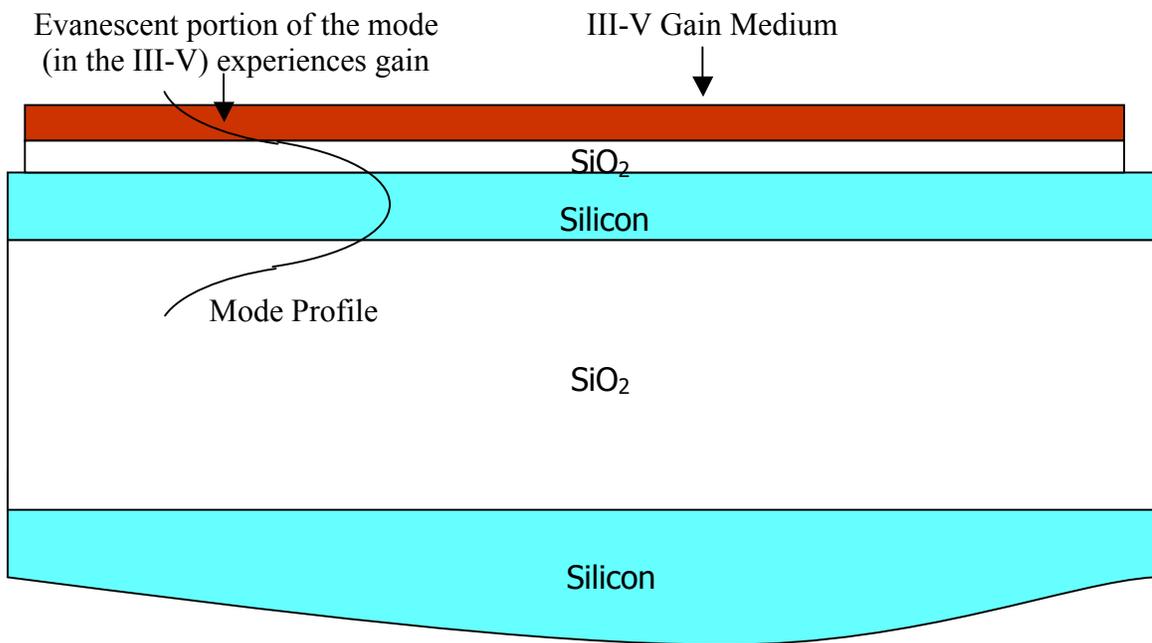


Figure 1-2: Design of an evanescent III-V/Silicon hybrid laser where only the tail of the cavity mode experiences gain

The motivation for taking this approach was the successful demonstration of cavities with extremely high quality factors ($>10^6$) on SOI^{21,22}. The implication of having such a high Q is that a very small amount of gain is required to overcome the losses in the

cavity and therefore the gain experienced by the decaying evanescent tail is enough to initiate lasing.

Even though this device still requires bonding of the III-V substrate, the constraints on that process are greatly alleviated due to the small footprint that the gain medium leaves on the Silicon chip and the fact that no major processing needs to be done on the III-V. The resultant reduction in cost coupled with the potential for electrical pumping, makes this geometry an attractive alternative to the devices that have been previously reported in the literature. Chapter 5 describes all these aspects of the device in much greater detail.

1.4 Organization of this dissertation

In this chapter we outlined the motivation for integrating photonic components on a Silicon chip in terms of improving efficiency and reducing costs. We also introduced the challenges involved in fabricating photo-detectors and lasers on Silicon and outlined our approaches to solving those problems. Chapter 2 looks at the fundamentals of the Ge/Si system while laying an emphasis on the materials science issues involved in the deposition of Ge on Si. An important point that is highlighted is the expected carrier transport mechanism at the Ge/Si junction and its effect on Ge film characterization and detector device design. Chapter 3 deals with the different techniques that we adopted to carry out the epitaxy of Ge on Si. The effect of various growth variables on the quality of the Ge layers is also outlined. The next chapter is devoted to the novel Photo-HFET structure that has been demonstrated. The design, fabrication, experimental realization

and analysis of the device are discussed in great detail. The evanescent Silicon laser is looked at in Chapter 5. The device is first treated comprehensively from a theoretical standpoint and then a detailed outline is provided for the experimental realization of the laser. Finally, our efforts towards the fabrication and demonstration of the device are discussed. Chapter 6 concludes the dissertation.

Chapter 2: Deposition and Characterization of Germanium Thin Films on Silicon: Theory

In order to optimize the deposition of Germanium films on Silicon substrates, the Ge/Si material system must be fully understood. In this chapter all aspects of the Ge epitaxy, including techniques for characterizing the thin films, are explored from a theoretical standpoint.

2.1 Issues with Ge growth on Silicon

One of the foremost requirements for successful epitaxial growth of one semiconductor on another is that the two semiconductors must have nearly identical lattice constants²³. Since the atoms of the deposited film align themselves to the atoms of the substrate, the presence of any lattice mismatch will create strain in the epitaxial film. As the thickness of the deposited film is increased, the overall strain in the material increases and a significant amount of strain-energy builds up. Finally a "critical" thickness^{23,24} is reached where it is energetically favorable for the material to release the strain by the creation of misfit dislocations and half-loops, which in turn can also lead to the formation of threading dislocations in the film. The strain energy is now concentrated around the dislocation lines only as opposed to being distributed throughout the material. Figure 2-1 shows a simple structural diagram that illustrates the formation of misfit dislocations.

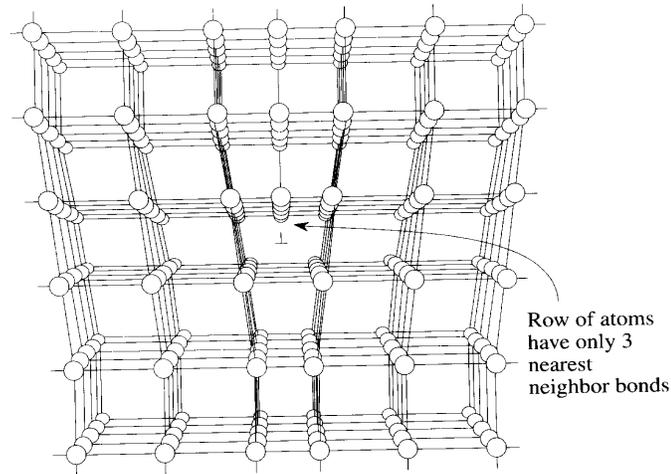


Figure 2-1: Misfit dislocation created by a lattice mismatch. The atoms in the dislocations have dangling bonds

The presence of a large number of defects in a semiconductor film can severely degrade device performance in any application. In the context of photo-detectors, the defects can drastically affect responsivity by acting as recombination centers for photo-carriers and thereby reducing internal quantum efficiency. Therefore, if epitaxial growth does involve lattice mismatch, the thickness of the deposited film should be kept below the critical thickness (t_c). Unfortunately in the case of Ge deposition on Si, there is a very significant mismatch between the two ($\sim 4\%$) and so the critical thickness, which is inversely proportional to the mismatch, is extremely small. Theoretical²⁵ and experimental²⁶ studies have shown that the value of t_c is about 10 \AA , i.e., the thickness of only 3 monolayers of Ge. Hence, an epitaxial Ge film of any practical thickness would have a high density of dislocations, which in turn, would introduce deep level states in

the Ge band-gap²⁷. These states are acceptor-like and as mentioned before they will act as recombination centers for photo-carriers thereby greatly degrading photo-detector performance.

2.2 Interfacial states and band alignment in Ge/Si

Since the lattice mismatch between Si and Ge is approximately 4% (Si: 5.431 Å, Ge: 5.658 Å), at the Ge/Si interface, roughly 8% of the Si surface atoms would have unfulfilled valancies (because Si has the smaller lattice constant). As the density of atoms at the Si surface is $\sim 10^{15}/\text{cm}^2$, the density of the surface states would be of the order of $10^{13}/\text{cm}^2$. Because these dangling bonds require electrons to fill their vacancies, they can be modeled as a sheet of acceptors at the interface^{28,29}, which in terms of energy, lie roughly in the middle of the Silicon band-gap. These acceptors can capture electrons from the two semiconductors, which in turn can cause surface depletion in both.

Another important factor affecting band alignment is the presence of a large number of defects in the epitaxial film. As discussed in the previous section, the dislocations in the Ge produce states in the band-gap that are acceptor-like. Due to this, the Ge film becomes p-type even if it is not intentionally doped, with the peak doping near the interface approaching $10^{18}/\text{cm}^3$ (the level of doping, which essentially corresponds to the density of defects, peaks at the Ge/Si interface and then reduces as the thickness of the film increases. This is because as the film gets thicker, there are interactions between dislocations with different Burger's vectors that can result in annihilation. Therefore the overall dependence of areal defect density (D) on thickness

(x) can roughly be expressed as: $D(x) = D(0)/[1+\lambda D(0) x]$, where λ is a characteristic length and $D(0)$ is the initial density. Empirically it has been found that the asymptotic value of the density of dislocations in the Ge film is between 10^7 and 10^8 cm^{-2} . For a more detailed study of Burger's vectors and dislocation kinetics please look at references 30,31).

Since the doping level in the Ge film is very high at the interface, the width of the depletion region in the Ge is insignificant. Therefore the Fermi level reaches its equilibrium position by causing depletion only in the Silicon, with the acceptors at the interface providing the holes needed to achieve charge balance. Figure 2-2 shows the resultant band alignment expected for p-Ge/n-Si. The Si doping shown approximately corresponds to the substrates used in our experiments ($\sim 10^{15}/\text{cm}^3$). The exact equilibrium position of the Fermi level is determined by all the factors discussed above: the density and energy levels of the Si surface states and the peak density of dislocations in the Ge. Its position at the Si surface is roughly expected to be 0.3 to 0.5 eV above the Si valence band.

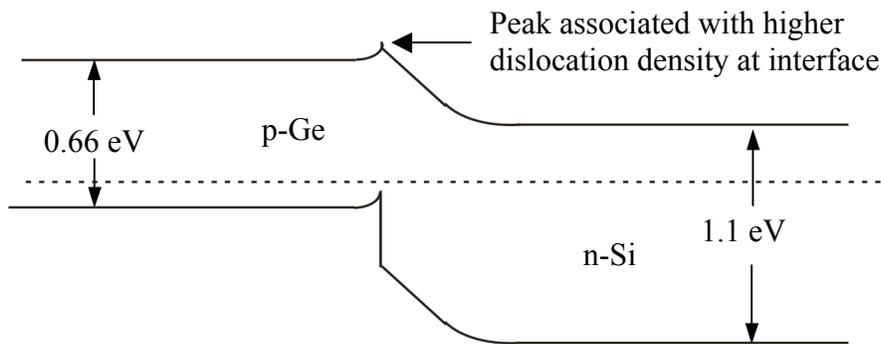


Figure 2-2: Expected band alignment in a p-Ge/n-Si structure. The depletion in the Si is attributed to surface states at the Ge/Si interface

2.3 Photo-carrier transport across Ge/Si junction

Since field assisted collection of photo-carriers is normally an extremely efficient process, the sensitivity and bandwidth of junction detectors is maximized if the depletion region is primarily in the active material³², which in the case of the Ge/Si system is the Germanium. Hence the normal band alignment in p-Ge/n-Si is clearly not beneficial to the overall performance of the detector. But, as was mentioned in section 2.2, the Ge was not intentionally doped and the p-doping arose primarily from the acceptor-like sites created by the dislocations. Since we cannot significantly reduce the density of dislocations using low temperature processing, the band alignment can't really be altered that much. Therefore, photo-carrier collection in any detector that we fabricate using the p-Ge/n-Si system, *will be limited to diffusion within the Ge*. This is an extremely important property of our photo-detectors and significantly affects their design and fabrication.

Rigorously, the diffusion photo-current (J_{diff}) in the Ge film is calculated by solving the continuity equation (including optical generation) for electron density (n_{Ge}) and then differentiating the result to give J_{diff} ,

$$J_{diff} = qD_n \left(\frac{\partial n_{Ge}}{\partial x} \right)_{x=0} \quad (2.1)$$

where, D_n is the diffusion coefficient for electrons in Germanium and $x=0$ refers to the edge of the Si depletion region (i.e., the Ge/Si interface). The solution for $n_{Ge}(x)$ would

be strongly dependent on the diffusion length (L_D) of electrons in the Ge film. L_D represents the average distance that an electron travels in the Germanium before it recombines. As there is no optical carrier generation within the Silicon, the photocurrent in the Si substrate is also equal to the right hand side of equation 2.1.

The current evaluation scheme described above is quite complete, but in order to better understand diffusion-dominated photo-response it is worth looking at a more simplistic model of the carrier collection process. Since the photo-carriers in the Ge diffuse a distance of L_D before they recombine, *only those carriers generated within one diffusion length of the Ge/Si interface will reach the Si depletion region and get collected.* In other words, if light is normally incident on the Ge film, only the portion of optical power absorbed within one L_D of the interface contributes to the photo-response. This model gives us a very useful tool in estimating diffusion length from responsivity, the details of which will be described below.

2.4 Extraction of diffusion length from responsivity measurements

From the discussion in the previous paragraph, it is evident that with a longer diffusion length, a larger percentage of the photo-carriers will be collected efficiently resulting in higher responsivity. Therefore if Germanium films can be produced with sufficiently long diffusion lengths then *high responsivity photo-detectors can be fabricated using efficient device designs.* Hence the initial focus of our experiments is on depositing Ge films on Si that exhibit the highest possible values for L_D . The second half

of this project deals with using those Ge films to create high responsivity photo-detectors (chapter 4).

In order to effectively characterize the Ge-on-Si films we need a reasonable method of measuring L_D . As was mentioned earlier, this can be accomplished by taking advantage of the fact that only the carriers generated within one diffusion length of the Ge/Si interface contribute to the photocurrent. Consider the simple photo-detector

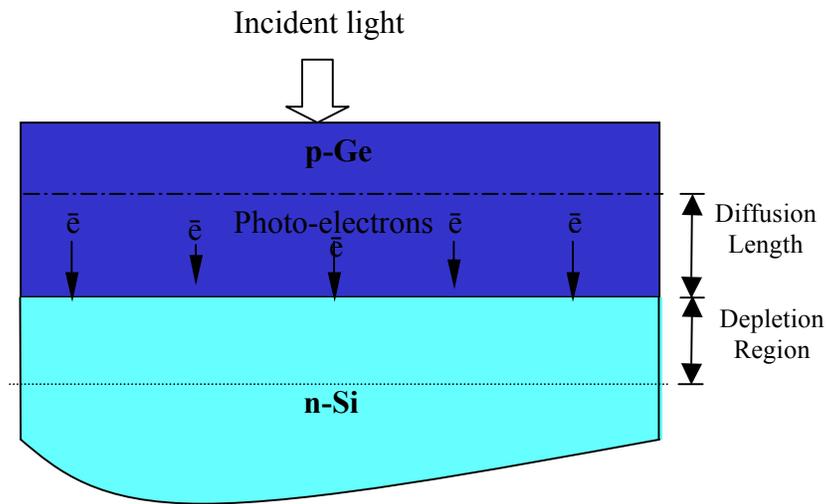


Figure 2-3: Photo-detector structure that can be used to estimate diffusion length of minority carriers in the Germanium

structure shown in Fig 2-3. Due to the aforementioned diffusion-dominated photo-carrier collection mechanism, the ratio of the light absorbed within one diffusion length of the interface to that absorbed by the entire Ge film is essentially equal to the internal quantum efficiency of the detector. Therefore,

$$\eta_i \approx \frac{(1 - e^{-\alpha L_n})}{(1 - e^{-\alpha d})} \quad (2.2)$$

where, α is the absorption co-efficient for Ge and d , the thickness of the Ge film. If we assume that L_D is very short, which is the case for highly dislocated Ge, the numerator reduces to αL_D . But it is also well known from simple p-n junction photo-detector theory³² that

$$\eta_i = \frac{R h c}{q \lambda (1 - R_{ref}) (1 - e^{-\alpha d})} \quad (2.3)$$

where, R_{ref} is the reflection coefficient for normal incidence at the Ge-air interface ($R_{ref} \sim 0.36$), h is the Planck's constant and c is the speed of light. Comparing these two equations, we get the following expression for diffusion length,

$$L_D = \frac{R h c}{\alpha q \lambda (1 - R_{ref})} \quad (2.4)$$

Therefore, if the responsivity of such a detector is measured at a wavelength λ then L_D can be calculated using eqn 2.4. This is the method that we utilize to measure L_D in the Ge films that are deposited using different growth techniques. The layout of the devices that we use for our measurements is also roughly similar to the structure shown in Fig 2-3. It is important to realize that the value of L_D given by eqn 2.4 is only a rough estimate. The exact diffusion length is extremely difficult to gauge because the varying density of dislocations in the Ge would ensure that the value of L_D doesn't necessarily stay constant over the entirety of the Ge film.

2.5 Epitaxial techniques for dislocation reduction

It was mentioned in the previous section that Ge films with long diffusion lengths are necessary to produce detectors with high responsivities. Therefore in order to facilitate the successful fabrication of Ge photo-detectors on Silicon, the number of dislocations in the epitaxial Ge must be minimized. This is because a high density of dislocations greatly increases the number of recombination centers in the Ge film which in turn drastically lowers L_D and carrier lifetime.

A lot of work has been reported in the literature on eliminating dislocations in Ge deposited on Si. The use of surfactants to reduce the surface energy of the substrate and thereby improve the quality of the epitaxial layer is well known. Both Hydrogen³³ and Antimony³⁴ have been employed successfully to create high quality Ge layers. Another approach is to partially remove dislocations after the growth has been completed by applying stress on the epitaxial film. The dislocation motion achieved by the stress³⁰ can result in annihilation through favorable interactions between dislocations or by driving the dislocations towards the edges of the film. An efficient method of creating stress is by alternately raising and lowering the temperature of the film, i.e., by subjecting it to "cyclic thermal annealing"³⁵ (the stress comes about due to the difference in thermal expansion coefficients of the substrate and the epitaxial film). This technique has been used very successfully with Ge, yielding films with very low threading dislocation densities^{6,7,27}. Another way of reducing the number of dislocations is by adopting selective area growth^{35,36} during the Ge epitaxy. Since deposition in a windowed area

brings the edges of the film closer, dislocations from a greater percentage of the growth area can now reach the edges and get annihilated. Therefore if the window is made small enough, this technique combined with cyclic thermal annealing can remove practically all the dislocations in the deposited film (Kimerling et al³⁶, demonstrate 10 μ m x 10 μ m mesas with no threading dislocations).

2.5.1 Limitations imposed by low temperature (<450 °C) processing

All the processes described in the previous paragraph are very effective in improving the quality of the epitaxial Ge film. Unfortunately none of these methods are back-end CMOS compatible because the temperatures involved in each of them are much higher than the ceiling temperature of 450 °C. The Ge deposition temperature in most cases is >500 °C and the cyclic annealing techniques involve alternating the film temperature between 750 and 900 °C. An additional problem lies in the preparation of Si wafers for epitaxy: in order to achieve the best possible results, the substrates must be cleaned using a high temperature process involving some form of oxide desorption (the Shiraki clean for example, which requires heating the sample to ~800 °C). Therefore, in order to stay compatible with back-end CMOS processing, new ways of improving the Ge films must be explored that limit the substrate cleaning and deposition temperatures to less than 450 °C.

In order to achieve a post-growth improvement in Ge quality an effective method to remove dislocations has to be introduced. But as mentioned earlier, any such dislocation reduction is extremely difficult to accomplish without resorting to very high

temperatures. Hence, our focus is primarily on improving the quality of the Silicon substrate prior to Ge deposition (which would in turn improve the crystallinity of the Ge film), using techniques that keep the temperature below 450 °C. It is well known that the nature of the substrate surface depends a lot on the chemicals used to do the cleaning, the annealing treatments used and also on the order in which these steps are executed. Our experiments explore these variables by changing the solutions used for the final step of the cleaning process and then adjusting the annealing conditions accordingly. It must be also be mentioned that in addition to using different cleaning mechanisms, we also studied the dependence of the Ge quality on the rate of deposition. The next chapter describes the exact growth processes in greater detail.

Chapter 3: Deposition and Characterization of Germanium Thin Films on Silicon: Experimental Results

As mentioned in the previous chapter, the goal of our measurements is to optimize the epitaxy of Ge on Si by tweaking various growth parameters. Clearly, in order to successfully fabricate the detectors that were introduced in chapter 1, Ge needs to be deposited on SOI. But due to the lack of availability of a large number of prime n-SOI substrates the epitaxial experiments are first carried out on bulk Si wafers. In the context of the different growth parameters that are explored (cleaning processes, growth rate etc) this switch from SOI to bulk Silicon has very little impact on the final conclusions that are drawn from the experiments. The Ge films that are used to fabricate the actual photo-detectors, are deposited on SOI substrates using the same process that is optimized for bulk Silicon.

3.1 Ge deposition details

Low temperature Ge deposition is done using both Molecular Beam Epitaxy (MBE, at 370 °C) and E-beam evaporation (at 300°C). As MBE is carried out at much lower background pressures (10^{-10} Torr *vis-à-vis* e-beam evaporation at 10^{-6} Torr), lower impurity incorporation and better quality Ge films are expected. Therefore the focus of our experiments is primarily on samples that are deposited using MBE. Additional parameters varied during the different depositions are the substrate cleaning method, the

Final Step of substrate preparation	Growth Rate	Growth Rate =	Growth Rate =	Growth Rate =
		2.5 Å/sec	1.0 Å/sec	0.2 Å/sec
Only HF clean		Sample M 4	Sample M 1	Sample M 5
HF clean + 450 °C Hydrogen Desorption (15 min)				Sample M 7
HF clean + 200 °C Hydrocarbon Desorption (50 min)				Sample M 9
HF clean + 200 °C Hydrocarbon Desorption (50 min) + 450 C Hydrogen Desorption (15 min)				Sample M 12
Piranha Clean + 800 °C Oxide Desorption				Control Sample M 3

Table 3-1 (a)

Final Step of substrate preparation	Growth Rate	Growth Rate =	Growth Rate =
		1.5 Å/sec	0.5 Å/sec
Only HF clean		Sample E 1	
HF clean + 200 °C Hydrocarbon Desorption (50 min)			Sample E 3

Table 3-1 (b)

Tables 3-1 (a) and (b): Processing conditions for (a) MBE grown and (b) E-beam evaporated Ge films. The Ge deposition is done at 370 °C in all the MBE samples and at 300 °C in all the E-beam samples. (The samples are numbered in the order in which they are grown; the M or E prefix indicates MBE and E-beam growth respectively)

annealing temperatures and the growth rate. Based on the substrate cleaning, there are two different classes of Ge samples grown: normal back-end CMOS compatible samples where the processing temperature never exceeded 450 °C and "control" samples where the cleaning is done at a higher temperature. The advantage of growing the control sample is that its results give us a rough idea of the performance that is lost by adhering to the 450 °C temperature limit.

Tables 3-1(a) and (b) list all the different configurations used for Ge deposition in our samples. The annealing schemes used in each case depend largely on the chemicals used to clean the Silicon. For the CMOS compatible depositions, the samples are cleaned using the modified Piranha method: first the substrate is dipped in Piranha ($\text{H}_2\text{O}_2 : \text{H}_2\text{SO}_4 = 3:5$) for 1 minute and then in aqueous HF solution ($\text{HF} : \text{H}_2\text{O} = 1:10$) for 1 minute. This procedure is then repeated 5 times. The combination of HF and Piranha helps remove most surface impurities as well as the thin layer of native SiO_2 present on the Si. The final dip in HF passivates the Si with H atoms and leads to a surface stable in air³⁷. The problem though is that the thin layer of H adatoms is believed to hinder the proper registration of Ge atoms with the Si substrate during the Ge deposition process. Therefore to obtain a Hydrogen-free surface we use a 450 °C *in-situ* anneal to try and desorb the H atoms. It must be mentioned that the annealing is done at 450 °C to make it compatible with the back-end processing ceiling temperature. Since complete Hydrogen desorption occurs at $>600^\circ\text{C}$ ³⁸, we expect some H atoms to still be present during epitaxy. In addition to the Hydrogen desorption anneal, some of the samples are also subjected to a long anneal at 200°C to remove hydrocarbons from the surface. Hydrocarbons are known

to cause problems in Si epitaxy and are usually removed by heating the substrate at moderate temperatures³⁹.

The cleaning and annealing steps for the control sample (M 3 in table 3-1a) are quite different from the other samples (for the E-beam evaporated samples, a control sample is not fabricated due to limitations imposed by the evaporator: the maximum temperature allowed in the machine is 300°C). Instead of finishing the clean by dipping in HF, sample M 3 is dipped in Piranha in the last step of the cleaning process. This ensures that the Si surface ends up being terminated by oxide when it is introduced into the MBE chamber. The oxide can then be desorbed using a high temperature *in-situ* anneal at 800°C. Following this oxide-removal technique results in the cleanest possible Si surface for epitaxy and therefore provides us with an adequate benchmark to gauge the quality of the films processed at lower temperatures. It can also be seen from the tables that in order to single out the effect of each annealing step on the Ge quality, the different anneals are not always used together.

The other details of the growth processes are as follows: the Si substrate in all the samples is n-type (ρ in the range of 2-6 Ωcm) with a (100) orientation and is single-side polished. The thickness of the Ge film in every sample is 200nm. In the MBE samples the first 190nms of Ge are doped lightly p-type using Boron, with the final 10 nm being doped p^+ to reduce the metal-Ge contact resistance. It was mentioned in chapter 2 that Ge deposited directly on Si exhibits p-type behavior due to the vast number of acceptor sites introduced by the misfit and threading dislocations. Therefore the Boron doping is not expected to make much difference in terms of the overall properties of the p-n junctions.

Finally, the base pressure in the CHA e-beam evaporator is 3×10^{-6} Torr while that in the Riber EVA 32 MBE chamber is 8×10^{-10} Torr.

3.2 Direct characterization of the Ge films

The Ge films deposited using the aforementioned growth schemes are directly characterized using various standard measurement techniques. An Atomic Force Microscope (AFM) is used to measure the surface roughness of the MBE deposited Ge films and the r.m.s roughness is found to be $\sim 1.5 \pm 0.4$ nm in all the samples.

During MBE growth, the deposition of Ge is monitored by Reflection High Energy Electron Diffraction (RHEED). In RHEED, initially the hydrogenated Si (100) surface is observed to be (1 X 1) reconstructed⁴⁰ at the growth temperature of 370 °C in all the samples (for the control sample, a (2 X 1) pattern is seen at ~ 800 °C, due to an oxide desorbed, cleaner Si surface). The RHEED pattern gradually changes from (1x1) to (2x1) during the Ge deposition process. Specifically, a transition from (1x1) to streaky (2x1) occurs at ~ 70 nm. Figure 3-1 shows a typical (2x1) RHEED pattern taken *after* the

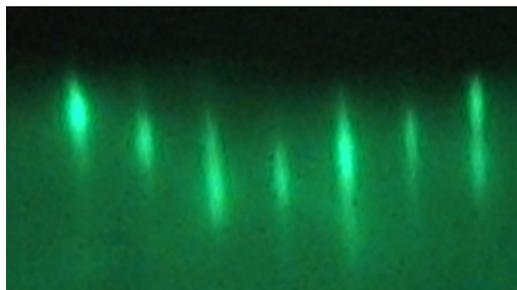


Figure 3-1: A typical (2x1) RHEED pattern after the completion of the Ge growth at 370 °C, indicating the formation of a single crystal film.

completion of the Ge deposition. The sharpness of the pattern developed indicates the quality of the Ge growth, and is found to be directly proportional to photo-detector performance.

The change in the RHEED pattern as the Ge growth progresses can be explained as follows: at low growth temperatures (< 400 °C), the diffusion of Ge adatoms is severely hindered and this in turn impedes the formation of a Ge wetting layer. Therefore, the Stranski-Krastanov mode⁴¹ is not followed and there is an initial period where the Ge growth is dominated by the formation of amorphous islands. Only after 50-70 nm of Ge deposition, do we see the emergence of a (2 X 1) RHEED pattern, indicative of crystalline material. At this stage, presumably, the effect of the surface roughness is diminished, and the misfit strain in the film has been accommodated by the formation of strain relieving defects. Ge diffusion is now facilitated and regular (2 X 1) dimers are formed.

The overall crystal quality and strain of the as-grown Ge/Si structures is analyzed by *ex situ* x-ray diffraction (θ - 2θ and rocking curve analysis) using the (004) reflections of the $\text{CuK}\alpha_1$ (1.5406 Å) line in a Bede3 diffractometer. The data from the XRD graphs (shown in Fig 3-2) can give us a rough estimate of the residual strain in the Ge films. Using the basic Bragg diffraction formula (with the lattice constant d replaced by $d/4$ owing to the fact that the (004) reflections are being used) we can derive the following expression for estimating the lattice parameter of Ge in the direction perpendicular to the plane of growth:

$$a_{\text{Ge}\perp} \sin(\theta + \Delta\theta) = a_{\text{Si}} \sin \theta = 2\lambda, \quad (3.1)$$

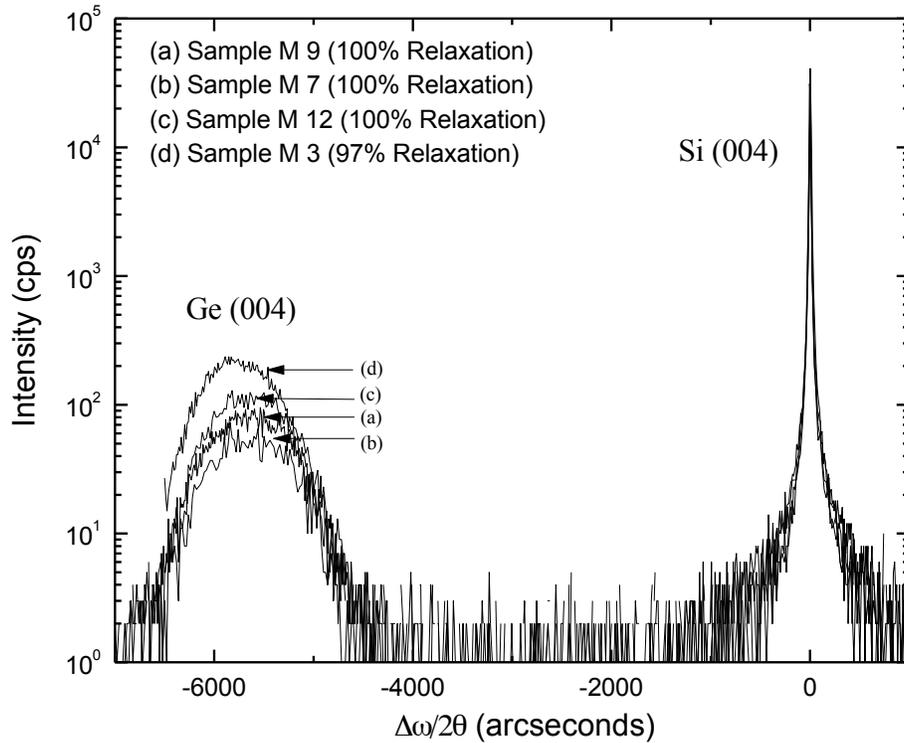


Figure 3-2: (004) double axis X-ray $\theta/2\theta$ curves of (a) Sample M 9 (b) Sample M 7 (c) Sample M 12 (d) Sample M 3. The peak position of the Si substrate is taken to be zero.

where θ is the Bragg angle for X-ray diffraction, $\Delta\theta$ is the angle between the Si (004) peak and Ge (004) peak and λ is the wavelength of the Cu $K\alpha_1$ radiation. The value that is obtained for a_{Ge} can then be compared with the lattice constant of bulk Ge to estimate the strain. From our calculations, we see that the Ge film in all the samples is almost fully relaxed. An important consequence of this is that the band-gap of the epitaxial Ge, and therefore its absorption coefficient, is identical to that of bulk Ge.

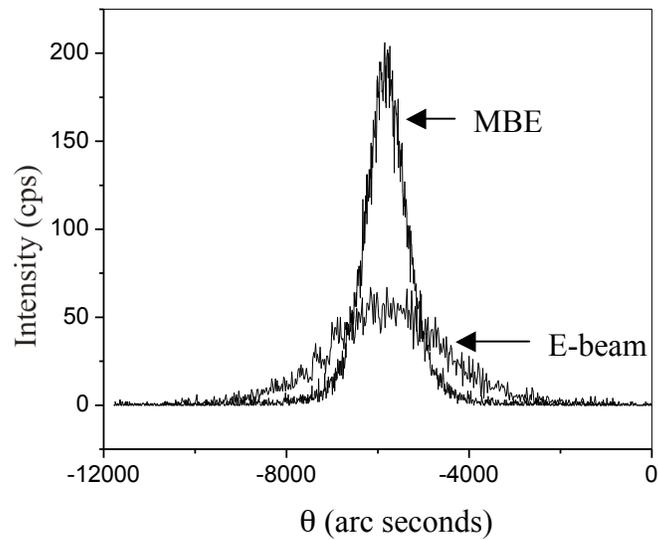


Figure 3-3: A comparison of the x-ray rocking curves of the MBE and e-beam films. The Full Width at Half Maxima (FWHM) is $\sim 0.3^\circ$ in the MBE sample, and $\sim 0.7^\circ$ in the electron-beam sample.

Fig 3-3 shows that XRD analysis can also be used for comparing MBE and e-beam evaporation. As the width of the rocking curve is inversely proportional to the crystalline quality of the deposited film, it is inferred from the figure that the MBE films are superior to the e-beam evaporated films, as expected.

Cross-sectional Transmission Electron Microscopy (TEM) is also used for analyzing the Ge films. Figure 3-4 shows the bright field TEM pictures for samples M 3, 7, 9 and 12. As expected, a dense network of dislocations and point defects is seen with the defect density peaking at the interface. Although it is impossible to count the number of dislocations at the interface precisely, we can get a rough estimate of the density of

dislocations at the middle of the Ge layer. From the figures we see that this density is $\sim 4 \times 10^{10} \text{ cm}^{-2}$ in all the samples.

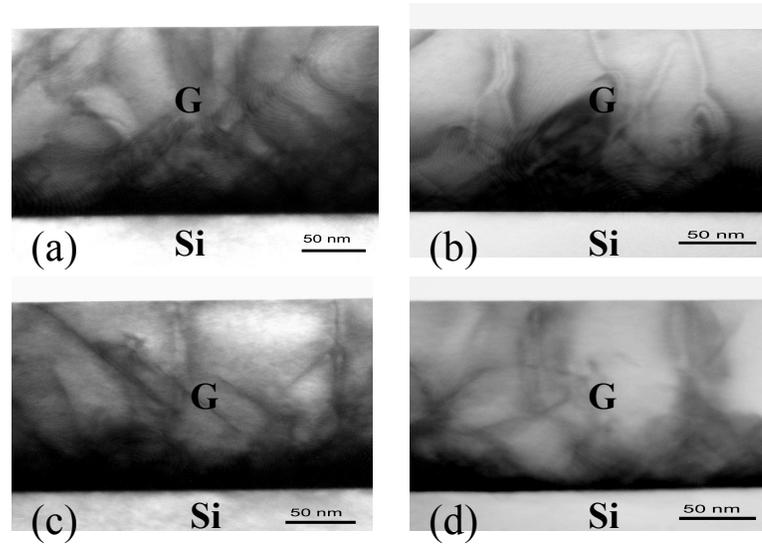


Figure 3-4: Bright field cross-sectional TEM images of the following samples (a) M 9 (b) M 7 (c) M 12 and (d) M 3

3.3 Device fabrication and experimental setup

The discussion in section 2.4 described the method we use to estimate electron diffusion length in p-Ge films deposited on Si. Figure 3-5 shows the exact structure of the photo-detector devices that we use to accomplish the measurement.

Experiments are done on both (a) large (5 mm X 5 mm) and (b) $400 \mu\text{m} \times 400 \mu\text{m}$ Ge mesa devices (the figure applies only to the mesas). Ge mesas are made using standard photolithography and wet etching: the etchant is a HF: H_2O_2 : Acetic acid

(1:2:30) mixture. Ohmic contacts are made to the n-Silicon and p-Germanium by e-beam evaporation at room temperature. Depositing Ti (15 nm)/Al (200 nm) on n-Si and Ag (300 nm) on Ge, gives the lowest contact resistance.

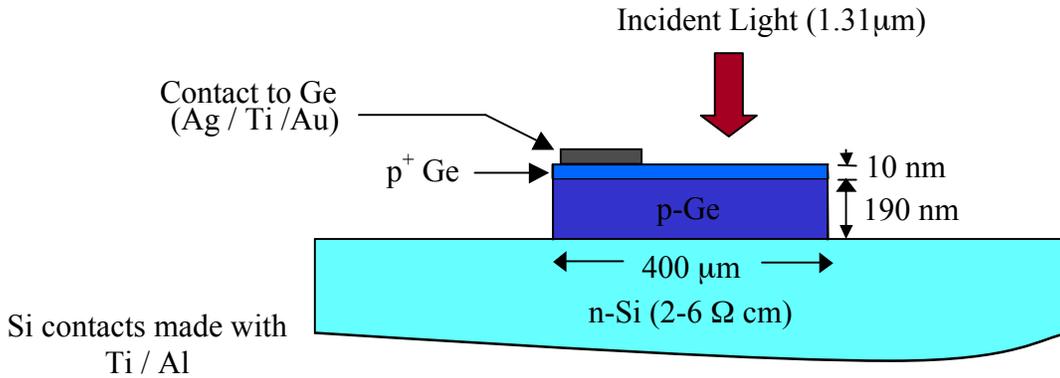


Figure 3-5: Structure of the mesa based p-Ge/n-Si detectors used in our electron diffusion length measurements.

The samples are illuminated by a 1310 nm laser (Thor labs mid-IR single mode laser diode), from the Germanium side in the 0-3 mW range. The responsivity (R) is measured using a standard AC lock-in technique: a fixed 9V DC supply is connected across the photo-detector and the reverse bias across the diode is varied by changing the resistance of a variable resistor (R_L) connected in series with the diode and supply. The laser light incident on the detector is modulated by an optical chopper and the output AC photo-voltage of the detector is measured using a DSP Lock-in amplifier (the lock-in gets its reference frequency from the chopper as well). The responsivity is given by:

$$R = \frac{2.22 \Delta V}{(R_j \parallel R_L) P_{inc}} \quad (3.2)$$

where, ΔV is the RMS value of the photo-voltage measured by the lock-in, R_j is the reverse resistance of the diode and P_{inc} is the incident optical power. There is a factor of 2.22 in the numerator because even though the output of the detector is a square wave (due to the chopping action), the lock-in only measures the RMS value at the fundamental frequency. Since the amplitude of the square wave is $\pi/2$ times the peak of the fundamental sine component, the actual photo-voltage generated by the detector is $\pi/2 \times \sqrt{2} \Delta V$, which is $\sim 2.22 \Delta V$. The presence of the parallel combination of resistors in the denominator can be understood by a straightforward Norton equivalent circuit. The photo-response is also measured from the difference of DC I-V curves, with *and* without light. The AC and DC measurements are in agreement for all the samples. Finally the responsivity value for each sample is converted to L_D using equation 2.4.

It is worth mentioning that the photo-response measurements done on small mesas are expected to be more accurate than those done on bulk devices because of "light trapping" effects in the Si substrate⁴². Since the Si is single-side polished the rough back surface of the substrate will consist of micro-planes that are arbitrarily oriented. Therefore even light falling at normal incidence on the back plane will encounter a wide range of incidence angles. As the critical angle for total internal reflection at the Si/air interface is only $\sim 16^\circ$, a significant portion of the incident light will be reflected without any Fresnel transmission loss. Hence in the case of our devices (shown in Fig 3-5), a major fraction of the light not absorbed by the Ge in the first pass will be reflected and

then re-absorbed by the Ge film. The resultant increase in responsivity introduces error into our diffusion length estimates because the formulation that we used to derive the dependence of L_D on responsivity (given by eqn 2.4) assumed that the incident light goes

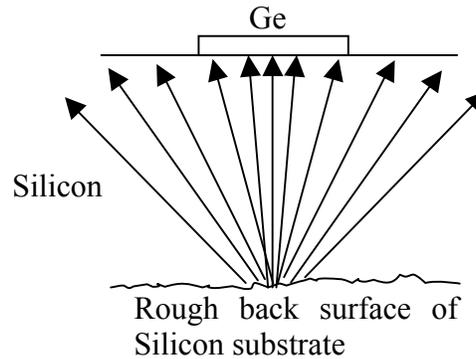


Figure 3-6: Total internal reflection at the back surface of a single side polished Si wafer. A small Ge mesa would avoid much of the back reflected light.

through the Ge only once. Since the back-reflected light re-enters the device at a wide range of angles (owing to the wide range of micro-plane orientation angles), the amount of light re-absorbed by the Ge depends on the solid angle that the Ge mesa subtends at the Si back plane (see Fig 3-6). If the mesa is small enough, most of the back-reflected light misses the Ge and doesn't get re-absorbed. As our devices use reasonably thick Si wafers ($\sim 600\mu\text{m}$), $400\mu\text{m}\times 400\mu\text{m}$ mesas are expected to see back-reflection based errors of not more than 10% (which are included in the error bars in the plots for L_D). Therefore all the diffusion length results listed in the next section are calculated from responsivity results for the $400\mu\text{m}\times 400\mu\text{m}$ mesas and not the bulk devices. In the context of back reflections it must also be noted that using double side polished Si wafers doesn't completely solve the problem because reflections from the smooth back surface can

create standing wave intensity patterns, which in turn can introduce severe errors in the diffusion length estimates.

3.4 Diffusion length measurements: Results and discussion

Figures 3-7 (a) and (b) give the electron diffusion lengths measured for samples processed under different conditions. It is seen from Fig 3-7(a) that L_D is a function of the desorption technique used. Although hydrocarbon desorption seems to be ineffective in enhancing the diffusion length in the present study, the hydrogen and oxide desorption steps do improve the deposited Ge film quality. The hydrogen desorption anneal at 450 °C increases the value of L_D (sample M 5 vs. M 7) from 15 nm to 25 nm; the 800 °C oxide removal (M 3) increases it from 15 nm to 60 nm, as could be expected. Our results on the e-beam evaporated Ge samples (with a substrate temperature of 300 °C, e.g., sample E3), gave results similar to those reported by Masini *et al*⁸.

It was mentioned earlier that even though complete H desorption occurs only at temperatures exceeding 600 °C, the temperature ceiling for hydrogen removal is set at 450 °C keeping in mind the limitations set on back-end CMOS processing. As a consequence some Hydrogen is still expected to be present on the Si surface during Ge epitaxy. It is seen from the control sample, that if the Si surface is perfectly clean then better quality Germanium can be obtained. Therefore we can conclude that if a method is devised to remove all the Hydrogen at low temperatures then diffusion lengths of at least 60 nm are achievable with processing temperatures restricted to less than 450 °C.

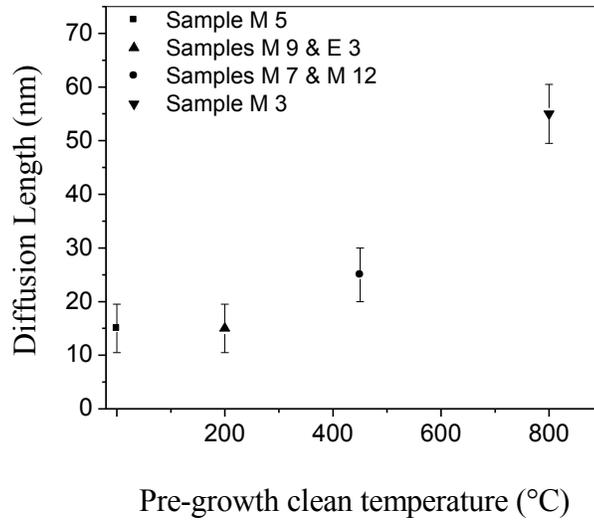


Fig 4-7 (a)

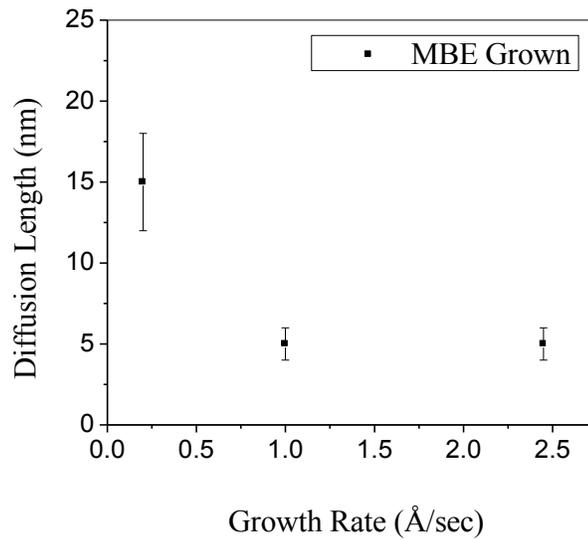


Fig 4-7 (b)

Figure 3-7: (a) Diffusion length as a function of pre-deposition anneal temperatures (b) Diffusion length as a function of growth rate. The points plotted in (b) represent samples grown with no prior annealing.

At this point it is instructive to compare the results of the diffusion length measurements and the structural analysis experiments (XRD, TEM) in terms of the differences seen between various samples. The improvement in the Ge films associated with an increase in diffusion length in some of the samples is not evident in either the X-ray analysis (fig 3-2) or the TEM images (fig 3-4). This discrepancy in the results can be attributed to the fact that an increase in L_D indicates *an improvement in the Ge quality very close to the Ge/Si interface (within ~20-30 nm)*, where the density of dislocations is too high to be effectively quantified by the structural analysis techniques used in our measurements. It must be noted that the increase in L_D most likely corresponds to a decrease in the number of interfacial point defects such as C precipitates and other inclusions typically associated with imperfectly cleaned surfaces.

Figure 3-7(b) shows the dependence of the electron diffusion length on growth rate. These results clearly illustrate that the quality of the epitaxial layer improves as the growth rate decreases. This is not surprising as a slower growth rate aids Ge adatom diffusion on the Si surface.

The DC current-voltage characteristics of a typical MBE grown p-Ge/n-Si diode are shown in Figure 3-8. The leakage current in all the MBE samples is in the range of 0.2-0.3 mA/cm² at -1 V reverse bias, which is among the lowest reported for Ge-Si based photo-detectors. The electron-beam evaporated samples, on the other hand, give higher leakage currents of about 1.6-1.8 mA/cm².

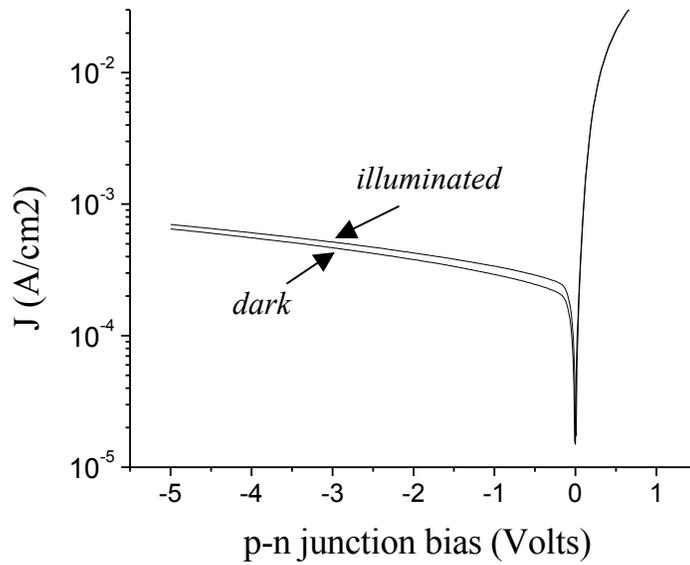


Figure 3-8: Semi-logarithmic plot of typical Current density vs. Voltage curves for the p-Ge/n-Si photodiodes, with (*illuminated*) and without (*dark*) light. All the devices exhibit nearly identical J/V curves

3.4.1 Leakage current modeling and analysis:

Since the leakage current in p-n junctions depends significantly on L_D , a reasonable method for testing the validity of our diffusion length estimates is to mathematically model the dark current measured in our devices. The simple model that we have constructed considers two mechanisms⁴³ that contribute to the leakage current in reverse bias: (1) the *generation current* produced by surface states at the p-n junction interface and (2) a *diffusion current* component arising from outside the space charge region in the Ge. Since the depletion region is almost completely in the Si (which has a very low intrinsic carrier concentration), the generation current within the space charge region is neglected.

Generation current: The leakage current from the Ge-Si interface, J_{int} , can be estimated as follows:

$$J_{\text{int}} \approx q n_{\text{Ge}} v_s \quad (3.3)$$

where, n_{Ge} is the minority carrier concentration in the Ge and v_s is the surface recombination velocity. n_{Ge} in our devices is $\sim 10^9 \text{ cm}^{-3}$ ($n_{\text{Ge}} = n_i^2/p_{\text{Ge}}$, where the intrinsic carrier concentration n_i is $\sim 10^{13}/\text{cm}^3$ and $p_{\text{Ge}} \approx 10^{17}/\text{cm}^3$). To get an idea of the maximum possible leakage current we assumed a reasonably high value of v_s ($=10^6 \text{ cm/s}$) Using these values we get J_{int} to be $\sim 0.15 \text{ mA/cm}^2$.

Diffusion current: An additional component of the leakage current arises from outside the depletion region. Minority carriers (electrons in p-Ge and holes in n-Si) diffuse to the edge of the space charge region, and are then swept by the reverse bias to the contacts. Due to the small band-gap of Ge, its minority carrier concentration is much higher than Si (by a factor of ~ 1000), and hence we only consider the diffusion leakage current on the Germanium side. The diffusion component ($J_{\text{diff},n}$) of the current in the p-Ge is estimated to be:

$$J_{\text{diff},n} \approx q \left(\frac{n_{\text{Ge}}}{\tau_n} \right) L_D \quad (3.4)$$

Here, τ_n is the minority carrier lifetime in the p-Ge. τ_n is estimated from the diffusion coefficient (D_n) and L_D . D_n can in turn be calculated from the mobility (μ_n) of electrons in Ge using the Einstein relation ($D_n/\mu_n = k_B T/q$). The expected value of μ_n for a doping level of 10^{17} cm^{-3} is $\sim 1000 \text{ cm}^2/\text{V s}$. But since the Ge is heavily dislocated we expect the

mobility to be at least a factor of 10 lower. Therefore using $\mu_n \sim 100 \text{ cm}^2/\text{V s}$ and $L_d \sim 25 \text{ nm}$, we obtain $\tau_n \sim 2.5 \text{ ps}$. Using this in the above equation we get $J_{\text{diff},n} \sim 0.15 \text{ mA/cm}^2$.

Combining $J_{\text{diff},n}$ and J_{int} , we see that the estimated reverse leakage current is $\sim 0.3 \text{ mA/cm}^2$, which agrees quite well with the observed values of $0.2\text{-}0.3 \text{ mA/cm}^2$ in the MBE samples. This model predicts the leakage in the e-beam samples to also be of the same order since they exhibit almost identical diffusion lengths. But, as mentioned earlier, the observed leakage in the e-beam samples is about an order of magnitude higher. This is most likely due to the poorer crystalline quality of the Ge in these samples (as evidenced by the X-ray rocking curves), which could make the electron mobility lower and the surface recombination velocity higher thereby increasing the overall leakage current.

In analyzing the reverse leakage it is worth repeating that the increase in density of dislocations near the Ge/Si interface appears as a small energy barrier in the Ge band diagram (it would be depicted by an up-slope in the conduction band). This barrier would impede the diffusion of electrons towards the depletion region^{7,27} and thereby affect any current flow in the device. Some of our results do show the presence of a defect-induced barrier at the interface: in particular, in some samples, a low defect density (indicated by higher photo-response) is accompanied by correspondingly high reverse leakage current. But in general these effects are not very significant and the departure from typical values of leakage current is quite small. So, even though the assumption of an energy barrier at the Ge/Si interface would make the leakage model more complete, we do not include a barrier in our analysis because its impact on the overall performance trends in our devices

is negligible. Since the leakage current is calculated using the observed values of L_D , the fact that the observed and estimated values of leakage are in agreement indicates that our estimates of diffusion length are quite reasonable.

3.5 Dependence of responsivity on wavelength

The spectral variation of the absorption coefficient for sample M3 is shown in Figure 3-9. The graph is obtained by measuring the responsivity of the device at wavelengths ranging from 1100 to 1550 nm (a white light source is used through a monochromator). The responsivity at each wavelength is then converted to the corresponding value for the absorption coefficient with the calibration done by assuming α to be 10^4 cm^{-1} at 1310nm.

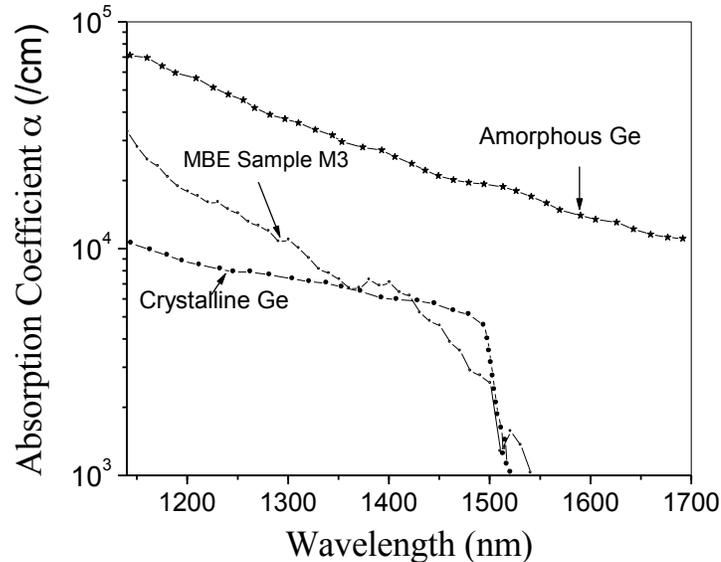


Figure 3-9: Measured absorption coefficient vs. wavelength for sample M 3 contrasted with the curves for single-crystal and amorphous Ge (data taken from Reference 7).

The curve shows that the absorption coefficient of the epitaxial Ge is roughly similar to that of bulk single crystal Ge, in the optical communications window of 1.3 – 1.55 μm . More importantly, the Ge film exhibits reasonable absorption even at 1550 nm and therefore efficient Ge/Si detectors, fabricated by the present methods, would be feasible at that wavelength. The graph for single crystal Ge exhibits a sharp transition at 1510 nm, which corresponds to the direct band-gap in Ge (0.8 eV). The lack of a similar transition in the measured curve (for sample M3) can be attributed to the poor crystallinity of the highly dislocated Ge-on-Si film.

3.6 Conclusions: Optimized deposition conditions

Since the goal of these experiments was to determine the optimum epitaxial conditions for Ge-on-Si, the important points mentioned in section 3.4 bear repeating here. The best quality Germanium films are obtained by finishing the wet clean with Piranha and then carrying out an *in-situ* high temperature oxide desorption at 800 °C. The estimated diffusion length for this technique is ~50-60nm which is at least a factor of 2 greater than any of the other Ge films. Even though the 800 °C temperature violates the back-end limit, the utility of that step in our experiments is worth revisiting. In addition to desorbing the native oxide from the Silicon, the anneal aids in the removal of impurities that might have adsorbed onto the surface during the wet processing preceding the growth. The incorporation of impurities onto the substrate is a more pronounced problem in a university clean-room environment and therefore the 800 °C bake becomes exceedingly significant.

In an actual foundry the problem of impurity adsorption would be much lower and hence the importance of a high temperature anneal would be diminished. Furthermore, the oxide on the Si substrate would be minimal if the Ge deposition is done following a process that also involves removing the native oxide (an ohmic contact deposition step for example). Therefore it might be a reasonable expectation than in an actual fab, Ge films deposited at 370 °C *without* a high temperature anneal will exhibit diffusion lengths in excess of 50nm.

Keeping this in mind, the Germanium films used in the proof-of-concept experiments on the novel Photo-HFET designs are grown with the 800 °C *in-situ* anneal in place. The improved results produced by the high temperature clean are also very beneficial in understanding the physical mechanisms responsible for the nature of the photo-response in those devices. The next chapter elaborates on the detectors that are fabricated using the optimized Ge-on-Si growth conditions described here.

Chapter 4: Design and Demonstration of a Ge/Si Photo-Heterojunction FET

As evidenced by the introduction given in the Chapter 1, Germanium-on-Silicon photo-detectors exhibiting very high speed and responsivity have been demonstrated using various growth techniques and device designs. In spite of the excellent performance of these devices, there are limits to the extent to which they can be miniaturized relative to conventional CMOS devices. In the context of optical integration, a photo-detector based on a MOSFET structure would represent a highly practical and uniquely scalable opto-electronic component. Here we propose and give proof-of-concept demonstrations of such a Ge photo-detector based on a transistor design that utilizes secondary photo-conductivity.

4.1 Design Principle

4.1.1 Charge separation at Ge/Si interface

The design of the device is derived from the band-alignment at the Germanium/Silicon heterojunction. Fig 4-1 depicts the expected alignment in our structures (the minor energy barrier in the Ge is not shown). Even though the reasons for why this exact band-bending occurs were discussed in detail in chapter 2, some of the main conclusions are worth repeating here. Due to the fact that most of the depletion region is in the Silicon, the collection of any photo-carriers created within the Ge film is

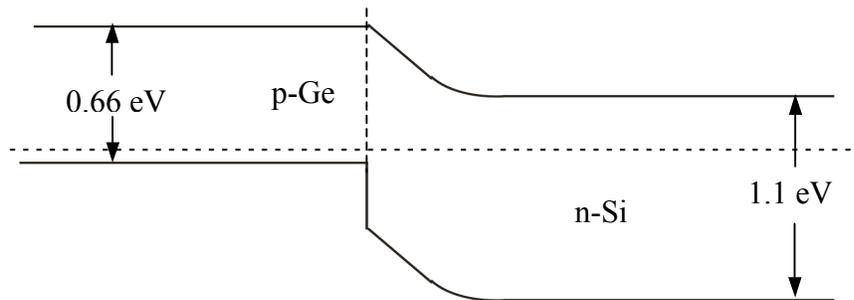


Figure 4-1: Expected band alignment in a p-Ge/n-Si structure.

dominated by diffusion. In addition, the photo-electrons are the carriers that are primarily collected, since the large valence band mismatch severely impedes the diffusion of photo-holes into the Silicon. Therefore in an open p-Ge/n-Si junction, the preferential collection

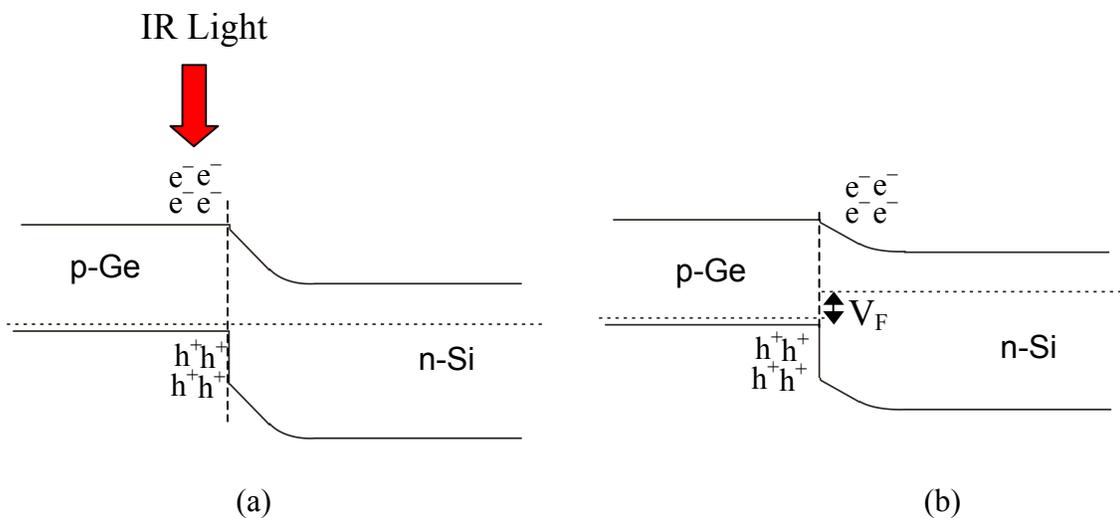


Fig 4-2: (a) Incident IR light creates electron-hole pairs in the Germanium (b) The electrons diffuse into the Silicon, creating charge separation. This forward biases the open junction by an amount V_F that depends on the amount of incident light.

of photo-electrons by diffusion would leave behind excess photo-holes in the Germanium. As illustrated in Fig 4-2, these holes will attract excess electrons in the Silicon, so as to maintain local charge neutrality. This creates a small open-circuit forward bias across the junction, the magnitude of which, is proportional to the number of carriers created and collected in the active region. This process can be referred to as secondary photo-conductivity because the excess photo-holes left behind in the Germanium can change the conductance of the Silicon.

4.1.2 Photo-detector structures based on the charge separation

An extremely practical method for utilizing this concept in a photo-detector is shown in Fig 4-3. A Germanium island is placed on the channel of a typical Silicon FET, replacing the gate that would normally control the conductance of the channel. An important point to note is that *no contact is placed on the Ge*.

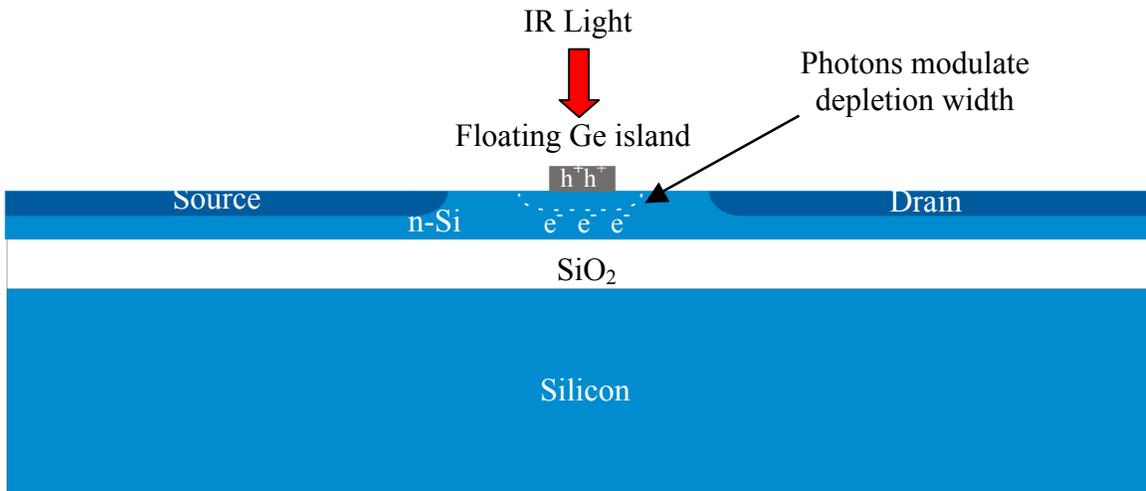


Fig 4-3: Cross-sectional diagram of the proposed Photo Heterojunction FET. A Ge island placed on an FET structure acts as the gate for the transistor. Since there is no contact on the Ge, it is floating.

The incident light modulates the channel depletion.

When light is incident on the Germanium, the Ge/Si junction gets forward biased due to the effect described above. This bias reduces the depletion in the Silicon and opens up the channel. This action emulates the operation of a regular ‘enhancement’ mode⁴⁴ transistor, where the channel is initially depleted. Clearly the ideal design for such a photo-detector also involves making a channel that is fully depleted initially and is then opened up by the light induced forward bias.

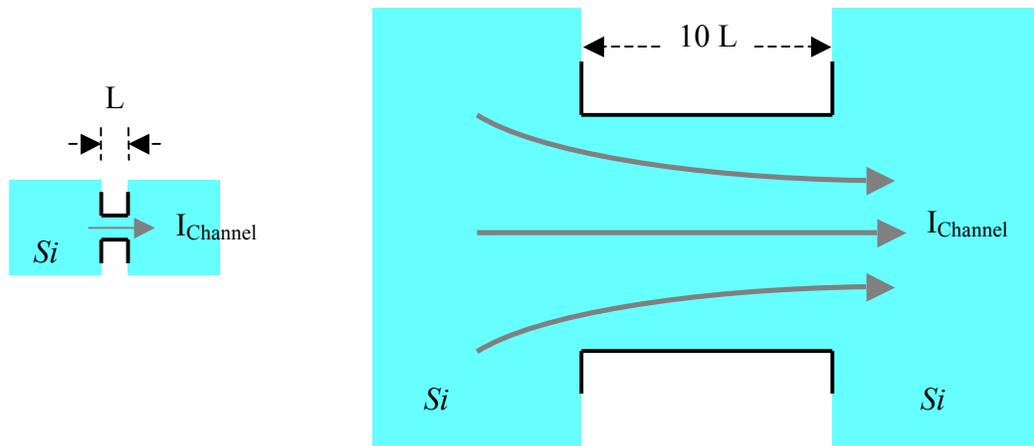


Fig 4-4: Top view of two devices with different areas but with the same W/L ratio. Due to the symmetric geometry, they will have roughly the same channel current. Also, light with the same *intensity* falling on the devices will cause an equal change in channel current.

The biggest advantage of this device stems from the fact that its sensitivity increases as its size decreases. This can be understood as follows: consider the top view schematic of two devices of different sizes shown in Fig 4-4, where the area of the bigger device is 100 times that of the smaller. Since the Si channel in both cases has the same Width/Length ratio and an equal amount of initial depletion, the channel resistance in

both devices will be roughly identical (in the figure, it is assumed to be 1 square of sheet resistance for the sake of simplicity). Now, to change the thickness of the depletion region by an amount Δt , $\sim N_D \Delta t$ *ionized donors per unit area* in the Silicon need to be neutralized and therefore $\sim N_D \Delta t$ *trapped photo-holes per unit area* would be required in both devices, where N_D is the n-doping of the Si channel. Since the number of trapped photo-holes per unit area simply depends on the intensity of the incident light we can reach the conclusion that the same *intensity* of light falling on both channels will roughly cause an equal change in depletion and therefore *an equal change in resistance* (ΔR). Hence if both detectors are biased to the same source-drain voltage, an equal intensity of light will cause an equal amount of ‘photo-current’, $V_{\text{Bias}}/\Delta R$ to flow in both cases. Therefore the smaller device, which is 100 times smaller in area, would require *100 times fewer photons to have approximately the same amount of photo-current*. Thus as these detectors are scaled down their sensitivity increases linearly with decreasing area. The advantages of this property are manifold and will be analyzed in greater detail in the final section of this chapter.

In light of the above discussion, it is clear that to exploit the favorable characteristics of the proposed Photo-HFETs, they should be made as small as possible. But an obvious drawback of the scaling is that the smaller devices will absorb much less light. This might not be a significant issue if the required photo-response is not very high, but the low quantum efficiency would still represent an optical power penalty in an actual communication system. One way to overcome that problem is to embed the device in an optical cavity. Fig 4-5 shows the schematic of such a structure: the incident light from the

waveguide makes multiple passes through the Germanium, thereby significantly increasing the amount of absorption. This design therefore, represents the ideal photo-detector based on the charge separation concept.

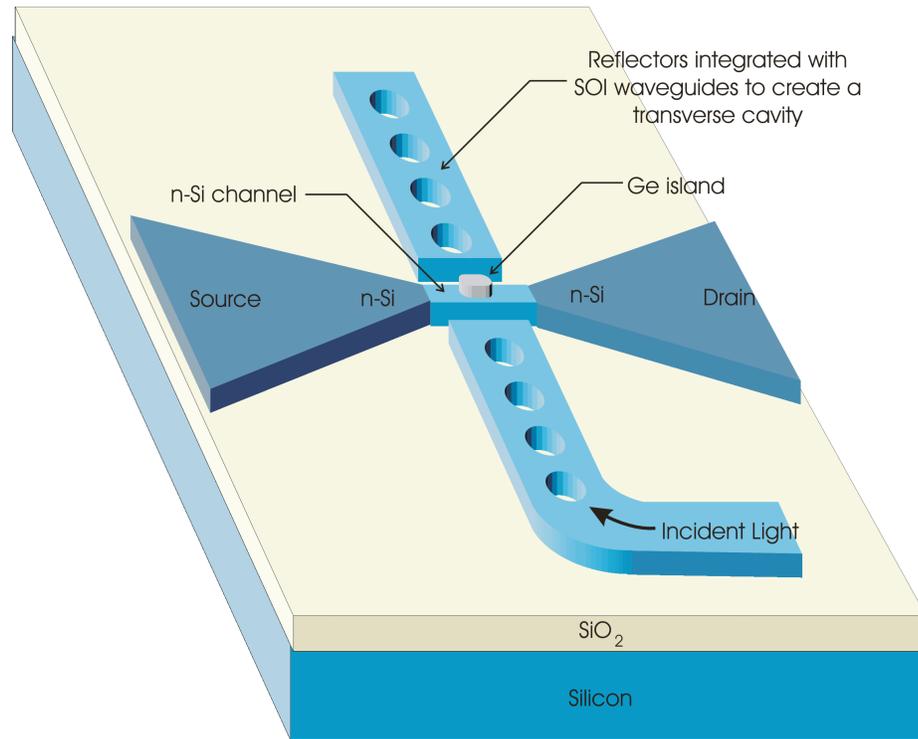


Fig 4-5: A Photo-HFET integrated with a resonator The detector can be scaled down to current MOSFET dimensions, with a transverse cavity compensating for the lower absorption.

4.2 Ge-on-SOI: Epitaxy and p-n junction characterization

The first step towards experimentally realizing the Photo-HFETs is the deposition of Ge films on SOI. Since the band alignment is most favorable in the p-Ge/n-Si configuration, the initial substrates need to be n-SOI wafers. But due to the lack of availability of n-SOI, the initial substrates consist of commercial p-SOI wafers from

SOITEC. These are then implanted with Phosphorus (dose: $5 \times 10^{11} / \text{cm}^2$ at 100 KeV) to make them n-type. The post-implantation anneal is done at 1150 °C and the final doping is found to be $\sim 2.5 \times 10^{16} \text{ cm}^{-3}$ with a small peak of $\sim 1 \times 10^{17} \text{ cm}^{-3}$ at the surface (through a SIMS analysis; Fig 4-6).

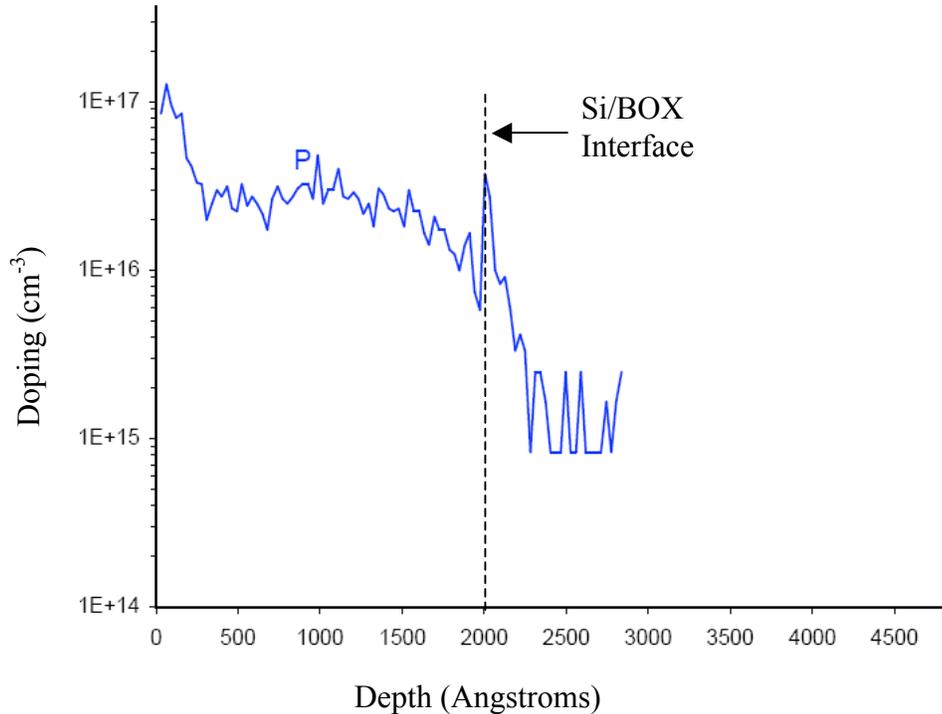


Fig 4-6: Results from a commercial SIMS analysis on the Phosphorus implanted SOI substrates (with Cs⁺ ions at 14.5KeV)

The Germanium growth is done using the optimum conditions discussed in sections 3.4 and 3.6. First a wet clean is done using HF and Piranha. This is then followed by an 800 °C *in-situ* anneal that desorbs the native oxide. Finally, the Ge epitaxy is carried out at 370 °C, which ensures the growth of a homogenous film with no islanding.

Following the technique to estimate diffusion lengths that was established in Chapter 2, simple p-Ge/n-SOI mesa based detectors are first fabricated (Fig 4-7). The Ge mesas are defined using photo-lithography and wet etching. The contacts to both the Germanium and Silicon are deposited using e-beam evaporation. The photo-response measurements are carried out with normally incident light at both 1.3 and 1.55 μm .

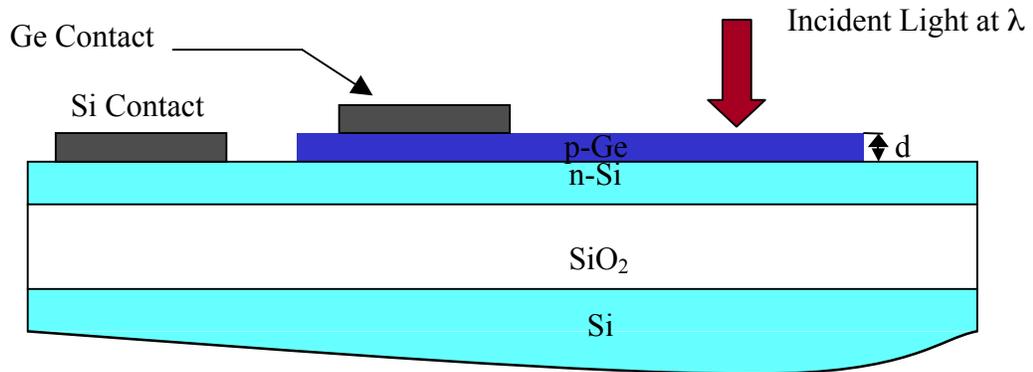
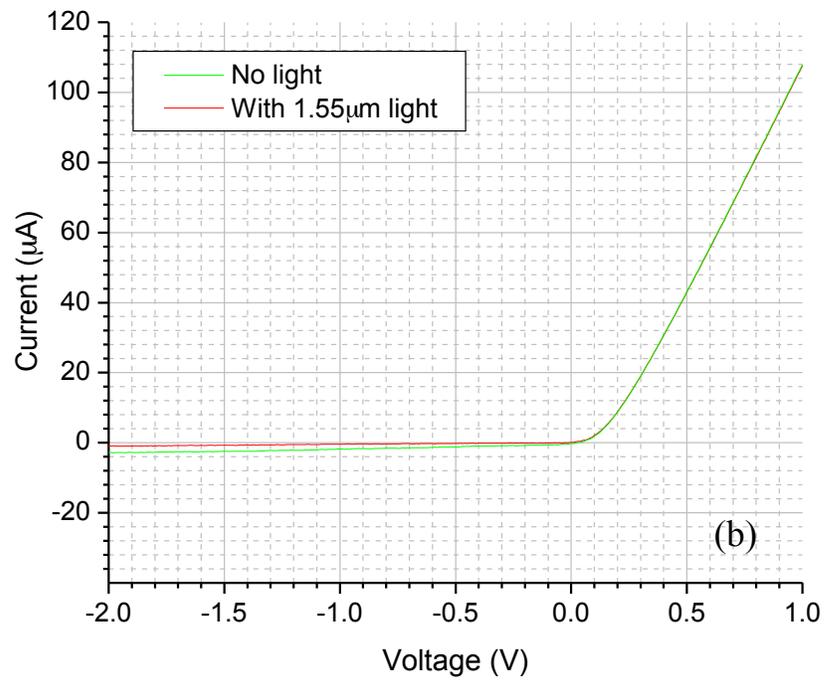
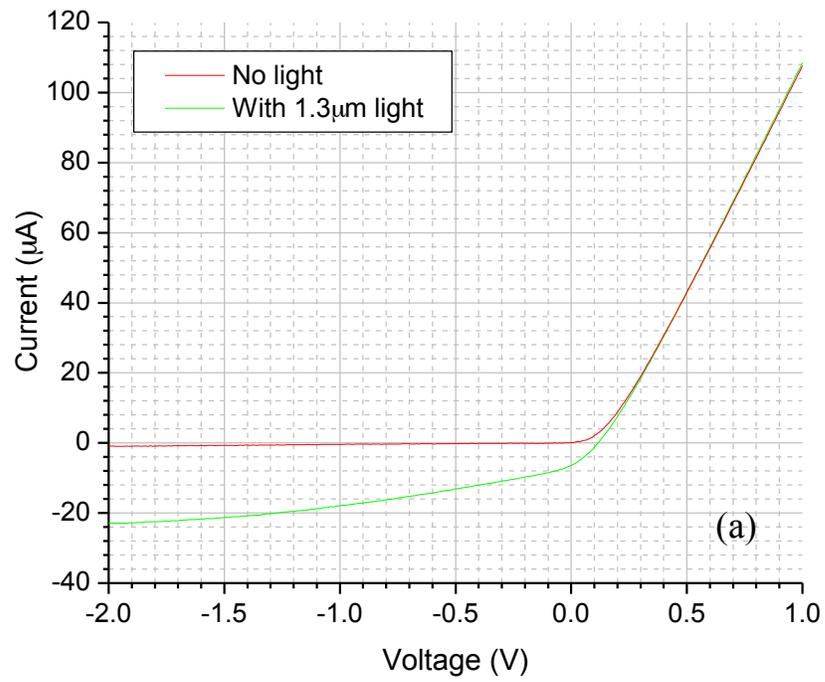


Fig 4-7: Cross-section of a simple p-n junction detector, whose responsivity is used to estimate the diffusion length of electrons in the Ge. The size of the Ge mesa is 400 μm x 400 μm .

4.2.1 Junction I-V curves

The I-V curves with and without light for both wavelengths are shown in Figs 4-8(a) and (b). It is clearly seen that the forward current is saturated by the series resistance in the diode circuit. This is to be expected as the top Silicon layer in the SOI is highly resistive. Therefore to get a more accurate picture of the diode I-V characteristics, the series resistance must be accounted for. Figs 4-8(c) and (d) show the I-V curves with *the actual voltage dropping across the p-n junction* on the x-axis. The corrected voltage is obtained by subtracting the potential drop across the series resistance for each bias point.



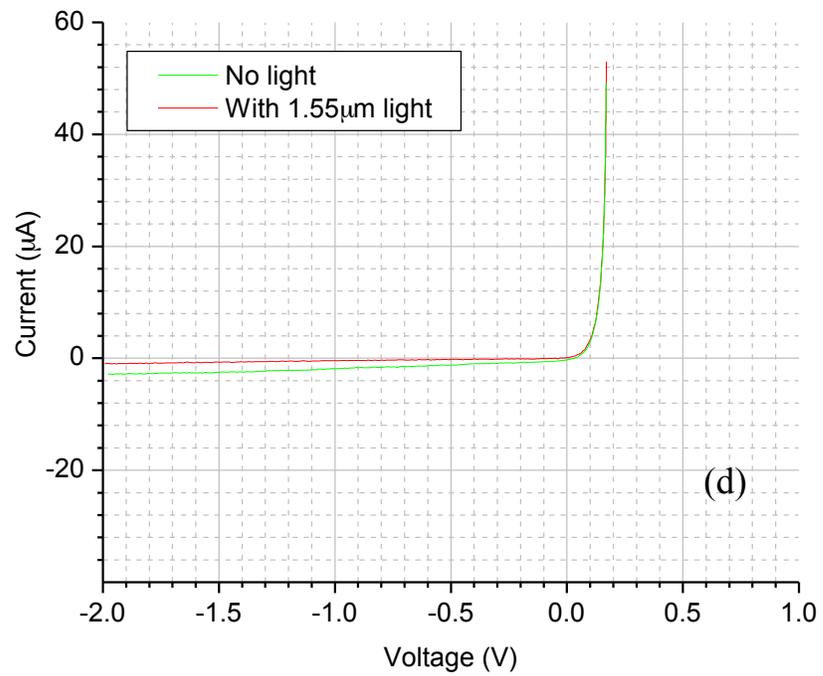
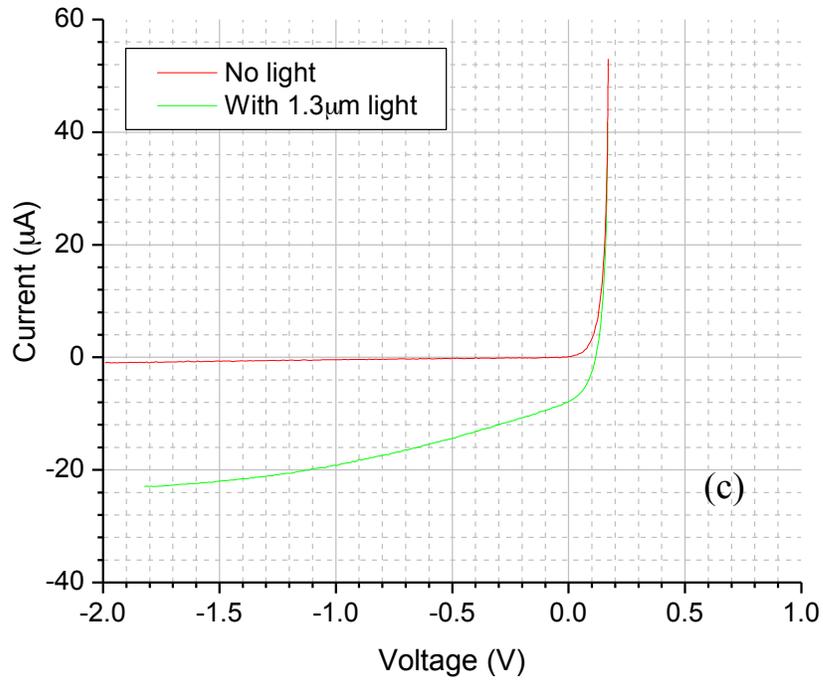


Fig 4-8: I-V curves with and without light for $400\mu\text{m} \times 400\mu\text{m}$ p-Ge mesas on n-SOI. Incident power is of the order of 1mW in all plots. (a) $1.3\mu\text{m}$ light (b) $1.55\mu\text{m}$ light (c) $1.3\mu\text{m}$ light; corrected for series resistance (d) $1.55\mu\text{m}$ light; corrected for series resistance.

The plots in 4-8(c) and (d) end at ~160mV of forward bias because that represents an approximate knee voltage beyond which the current increases extremely rapidly. Another interesting aspect of the curves is the increase in photo-response as the reverse bias is increased. This can be attributed to a lowering of the slight energy barrier on the Germanium side of the Ge/Si interface, which in turn facilitates the collection of the photo-electrons.

Fitting the IV curves to the diode equation is crucial for analyzing the subsequent devices because it gives us an excellent tool to parameterize the p-Ge/n-Si junction. The following expression accurately fits the *dark* plots (Fig 4-9 shows the fit):

$$I_{Diode} = 1.8 \times 10^{-7} \left(e^{\frac{qV}{1.27 k_B T}} - 1 \right) \text{ Amps} \quad (4.1[a])$$

i.e., the dark current is ~0.18 μA and the diode quality factor is ~1.27.

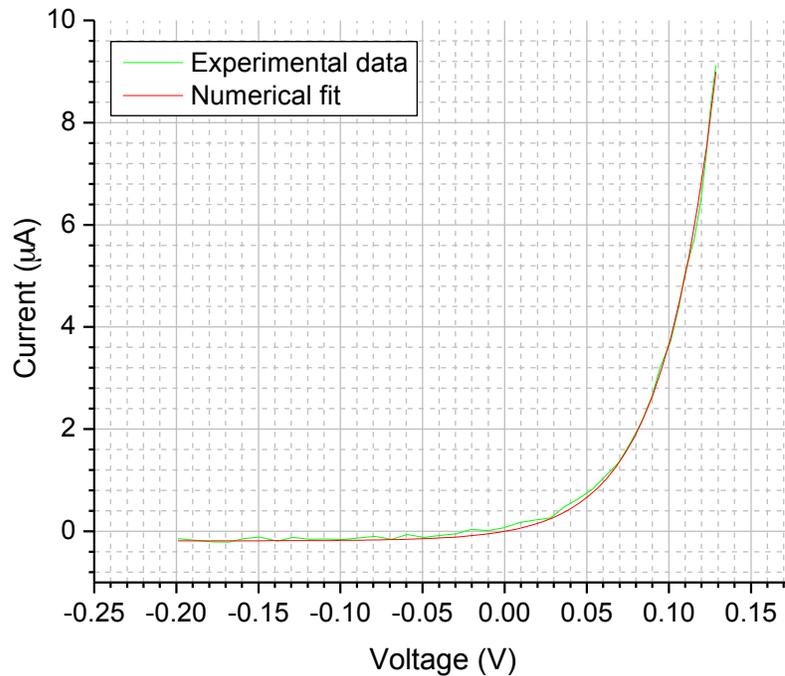


Fig 4-9: Numerical fit of the diode equation to the experimental I-V curve for the Ge/Si junction.

The dark current is specific to the 400 μm x 400 μm mesas and will scale down with area for the smaller devices. Hence normalizing to the area of the junction we get:

$$J_{Diode} = 1.125 \times 10^{-4} \left(e^{\frac{qV}{1.27 k_B T}} - 1 \right) \quad (4.1[b])$$

where, J_{Diode} is in Amps/cm². It is worth stressing again that these extracted parameters are extremely crucial in gaining a proper understanding of the mechanism that governs the behavior of the Photo-HFET detectors. The analysis in the next few sections treats those aspects in greater detail.

4.2.2 Diffusion length estimates

From the responsivity at zero bias, using expression 2.4, the diffusion length is estimated to be ~8 to 10 nm (assuming absorption coefficients of ~10⁴ cm⁻¹ and 400 cm⁻¹ respectively at 1.3 and 1.55 μm). This measured value of L_D is clearly much less than that observed for Ge films on bulk Silicon substrates. The implantation of the SOI substrates to make them p-type is thought to play an important role in degrading the quality of the deposited Ge layers. In any case, the effect of the lower diffusion lengths on the utility of the detectors is crucial and is analyzed in the last part of this chapter.

4.3 Initial devices: Fabrication and results

In order to verify the ideas described in section 4.1, basic proof of concept devices are first fabricated (Fig 4-10 shows the schematic diagram of the structures). The devices are made without highly doped source and drain regions in order to keep the process flow straightforward and therefore four contact pads are made on the Silicon in order to

facilitate 4-point probe measurements⁴⁵ that would overcome the problem of poor Ohmic contact resistance. The aim is to detect a change in channel conductance when infra-red light is incident on the gate.

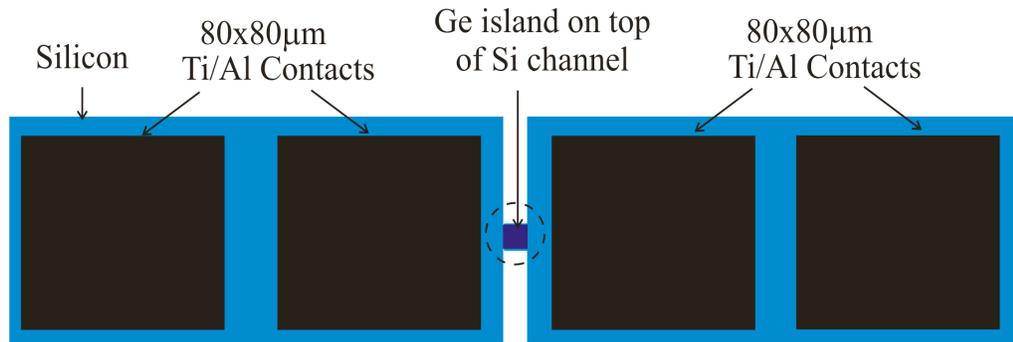


Fig 4-10: Top view of the initial proof of concept devices. A 4-point configuration is needed because proper ohmic contacts are not fabricated. The channel is the square of sheet conductance at the center.

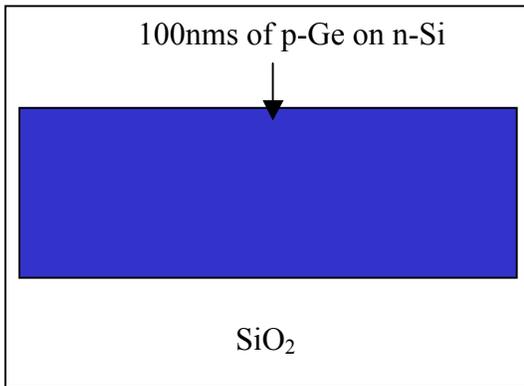
An important feature of the devices is that since the SOI substrate, which is only 200nm thick, has very low doping ($\sim 2.5 \times 10^{16} \text{ cm}^{-3}$) it is highly depleted at the p-Ge/n-Si interface. Any voltage applied across the source-drain contacts that raises the potential of the Silicon, further reverse biases this junction and the transistor enters the ‘saturation’ regime rather quickly. The saturation is just like the one seen in a normal FET⁴⁴, where the end of the n-Si channel closer to the positive terminal of the bias voltage reaches full depletion and the channel current stops rising proportionately with the source-drain voltage. Additionally the low doping underneath the contacts means that a similar effect

is seen at the Metal-Silicon interface, thereby further increasing the total channel resistance measured in the absence of light.

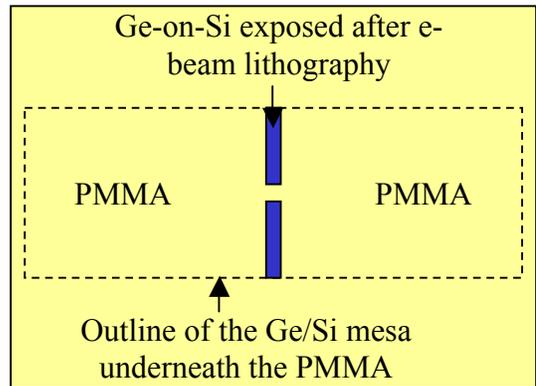
4.3.1 Fabrication

Fig 4-11 illustrates the various stages of the fabrication (all the figures depict the *top view*). The initial substrate is an n-SOI piece covered with a 100nm thick p-Ge film. The following steps are then undertaken:

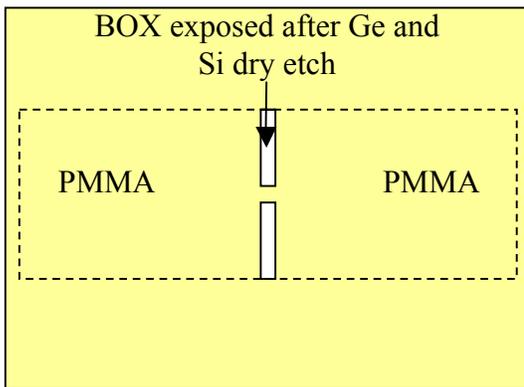
1. A large Ge-on-Si mesa ($\sim 600\mu\text{m} \times 100\mu\text{m}$) is first defined using photo-lithography and dry etching. Germanium and Silicon are etched all the way to the buried oxide in a reactive ion etcher using CF_4 and a $\text{SF}_6/\text{C}_4\text{F}_8$ mixture respectively. The size of this mesa is determined by the size of the contact pads needed for wire bonding.
2. A thin layer of Poly-methyl Methacrylate (PMMA) e-beam resist is spun on the sample and is patterned using electron-beam lithography to define the channel width and length. The channel is always made as a square; its dimensions are varied from $250\text{nm} \times 250\text{nm}$ to $4\mu\text{m} \times 4\mu\text{m}$. E-beam lithography is used since the processing is done in a university clean room. State of the art photo-lithography that is found in industrial foundries can easily accomplish the feature sizes that are used in these devices.
3. Germanium and Silicon are dry etched all the way to the buried oxide using the same etch process as in step 1.
4. The PMMA is removed in acetone.
5. A thin film of UVN-30, which behaves like a *negative* e-beam resist, is spun on.



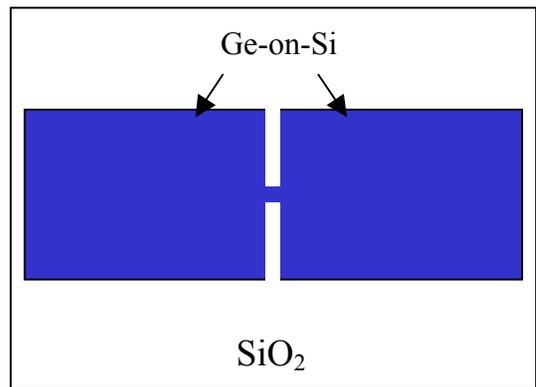
STEP 1



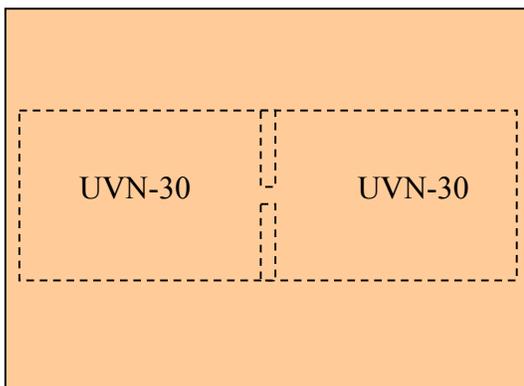
STEP 2



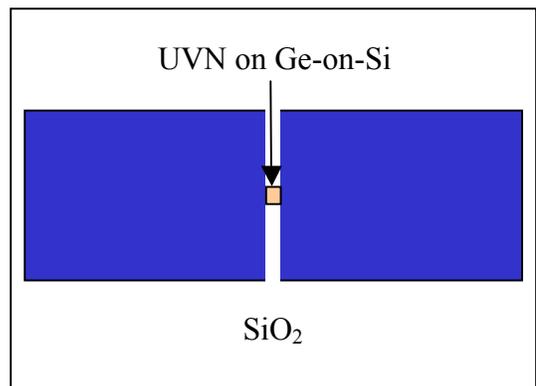
STEP 3



STEP 4



STEP 5



STEP 6

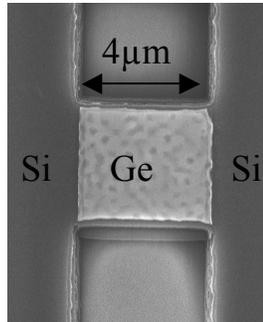


Fig 4-12: SEM image of a device with a channel length and width of $\sim 4\mu\text{m}$.

4.3.2 Experimental Results

The photo-detectors are characterized with normally incident light at $1.55\mu\text{m}$ that is focused onto the Germanium using a microscope objective. The beam waist at the focal

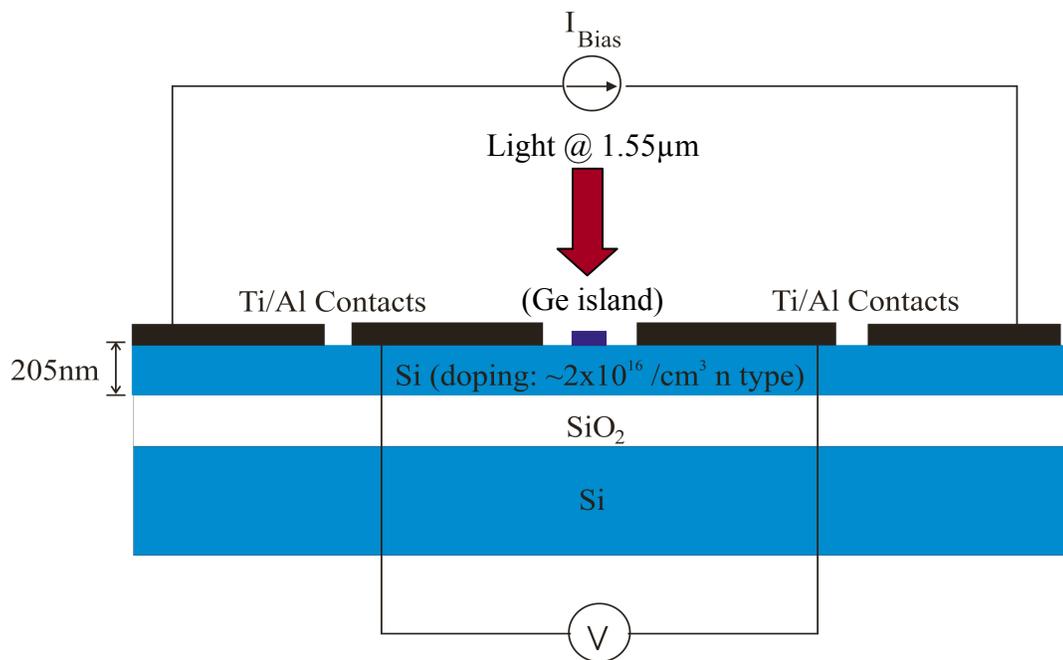


Fig 4-13: 4-point probe setup to measure the change in channel conductance caused by light.

point is $\sim 8\mu\text{m}$. The experiment is carried out as a standard 4-point probe measurement (Fig 4-13), where a constant current source is connected to the outer terminals and the voltage is monitored across the inner electrodes. The response of the device is essentially the change in channel conductance, $\Delta V/I_{\text{Bias}}$, where ΔV is the change in the voltage drop across the inner contacts when light is incident on the gate.

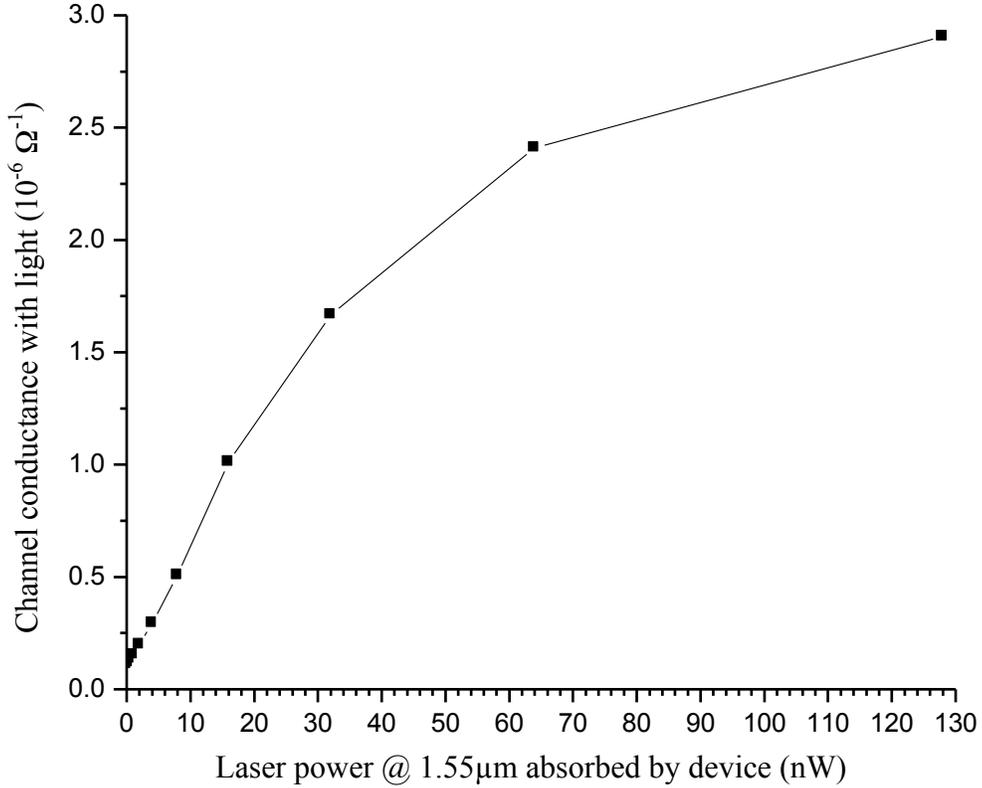


Fig 4-14: Change in conductance of a $1\mu\text{m}$ channel with increase in absorbed optical power

Fig 4-14 shows the change in conductance for different incident power levels for a device that has a $1\mu\text{m}$ long channel. The x-axis represents the total light absorbed by the 100 nm thick Ge gate and is given by

$$P_{Abs} = P_{in}(1 - R_{ref})(1 - e^{-\alpha t}) \frac{A_{Gate}}{A_{Spot}} \quad (4.2)$$

where P_{in} is the incident power, R_{ref} is the reflection coefficient at the Ge/air interface (~ 0.4), α is the absorption coefficient at $1.55 \mu\text{m}$ (400 cm^{-1}), t is the thickness of the Ge film (100nm), A_{gate} is the area of the gate ($1 \mu\text{m}^2$) and A_{spot} is the area of the beam spot ($\sim 50 \mu\text{m}^2$). As mentioned earlier, the low doping of the Silicon and the lack of proper source and drain contacts means that the channel resistance is extremely high without the light. For the device shown in Fig 4-14, the ‘dark’ resistance is of the order of 10 M Ω . Therefore the bias current is set to only 50 nA, so that the voltage across the device is roughly 0.5 V.

As can be seen from the plot, the device is extremely sensitive to the incident light. Owing to the high initial resistance it is seen that *absorbing $\sim 100 \text{ nW}$ of power increases the conductance of the channel by a factor of more than 25.*

It is worth repeating that the light that actually contributes to switching the conductance is the amount absorbed within one diffusion length of the Ge/Si interface. That amount is approximately a tenth of the total light absorbed because L_D is about 8 to 10 nm and the total thickness of the Ge is 100nm. Therefore if the thickness of the Ge gate was only 10nm, the strength of the effect would still be roughly the same. Hence in that regard the actual sensitivity of the device is *ten times higher* than what can be inferred from Fig 4-14.

4.3.3 Analysis

It was mentioned in section 4.1.1 that light incident on the open p-Ge/n-Si junction forward biases it. The expected change in channel resistance due to the light can be modeled by estimating the value of this forward bias. This is done as follows: *the*

magnitude of the bias will be such that it results in the flow of a forward current that perfectly balances the 'photo-current' created by the collection of the photo-electrons, thereby maintaining zero overall current across the open junction. Once the value of the forward bias is known it can be gauged whether the reduction in depletion in the Si caused by it is sufficient to significantly change the channel conductance.

As was established in chapter 2, the photo-current in the diode is:

$$I_{Photo} = \frac{qP_{in}(1 - R_{ref})(1 - e^{-\alpha L_D})}{(hc / \lambda)} \frac{A_{Gate}}{A_{Spot}} \quad (4.3)$$

where, the various variables have the same definitions that were used in eqn 4.2.

Comparing eqn 4.3 to eqn 4.2, we get:

$$I_{Photo} = \frac{qP_{Abs}}{(hc / \lambda)} \frac{(1 - e^{-\alpha L_D})}{(1 - e^{-\alpha t})} \approx \frac{qP_{Abs}}{(hc / \lambda)} \frac{L_D}{t} \quad (4.4)$$

Since $L_D \approx 10\text{nm}$ and $t = 100\text{nm}$., I_{Photo} (in Amps) $\approx 0.125 P_{Abs}$ (in Watts). Using this relation, the behavior exhibited by the device in Fig 4-14 can now be analyzed. It is seen from that plot that the change in conductance begins to saturate at absorbed power levels of $\sim 100\text{nW}$. According to eqn 4.4, this corresponds to a photo-current of the order of $\sim 10\text{nA}$. Therefore the forward bias created across the junction due to this photo-current will be such that $\sim 10\text{nA}$ of forward current flows. Plugging this into the diode equation for a $1\mu\text{m} \times 1\mu\text{m}$ p-Ge/n-Si junction (eqn 4.1[b]), we see that the *required bias is 0.3 V*.

An inspection of the band diagram of the p⁺-Ge/n-Si interface depicted in Fig 4-1 shows that the built-in potential at the junction is ~ 400 mV for the measured doping level

in the Silicon. Therefore the 0.3V forward bias created by the incident light is clearly enough to almost completely flatten the bands. This is what is seen in the plot in Fig 4-14, where a saturation in the increase in conductance occurs as the induced forward bias approaches 300 mV. The overall resistance of the channel does not fall below $\sim 500 \text{ K}\Omega$ however because of the high parasitic resistance associated with the contacts and the depleted Si channel not covered by the Germanium.

In analyzing the sensitivity of the device it should be stressed that the swing in channel conductance from dark to light is exaggerated by the large dark resistance of the channel. As mentioned earlier, the massive initial resistance (which is more than an order of magnitude higher than what is expected) can largely be attributed to the poor quality of the ohmic contacts to the Silicon that leave the channel almost completely depleted.

4.3.4 Conclusion

The simple proof of concept experiment described here shows that detectors based on the Photo-HFET principle can be exceptionally sensitive, especially for CW measurements. The absolute responsivity of the device is also very high since only $\sim 100 \text{ nW}$ of absorbed light changes the channel current by about $1 \mu\text{A}$ giving a photo-response of 10 A/W . These numbers are encouraging and give the indication that actual high speed detectors for optical communication systems based on this design principle can also be very effective.

The main obstacle in the realization of high speed detectors is the excess resistance coming from the areas of the Si channel not under the Ge gate. The

experiments described in the next section tackle that problem and show very promising bandwidth results.

4.4 Self-aligned detectors: Fabrication

4.4.1 Solid Phase Epitaxy

As mentioned in the previous section, the next step toward making practical high speed detectors is to reduce the contact and parasitic resistance associated with the channel. Therefore the required device structure involves creating highly n-type source and drain regions while leaving the channel with a lower doping (Fig 4-15). Typically the highly doped regions would be fabricated using high dose ion implantation followed by a high temperature anneal to repair the damage and activate the dopants. The annealing temperature is normally in excess of $1000\text{ }^{\circ}\text{C}^{47}$, which is well above the melting point of Germanium (940°C)²⁹. Additionally, the inter-mixing between the Ge and Si is expected to be very pronounced in that temperature range.

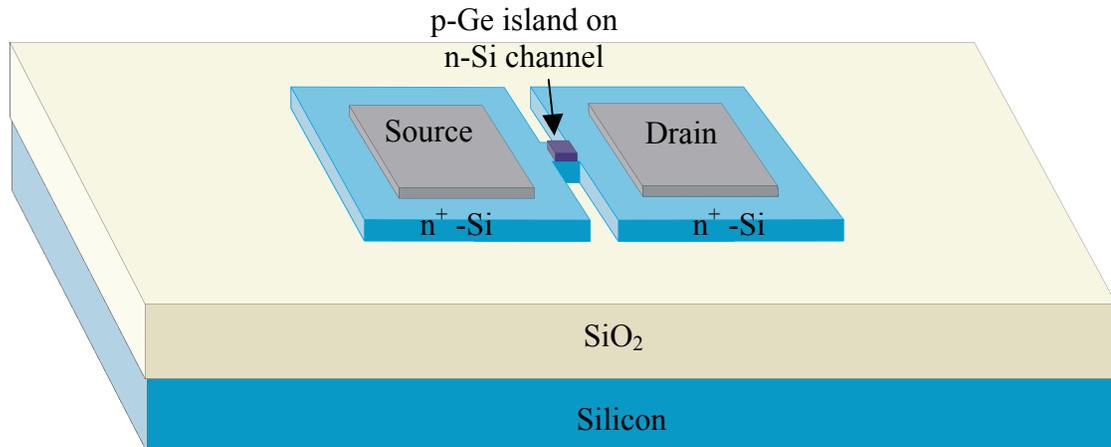


Fig 4-15: Schematic of a self-aligned device, where the top n-Silicon is heavily doped everywhere except underneath the Ge gate.

To avoid these problems, ideally the Ge epitaxy should be carried out after the anneal. Unfortunately the implant and the subsequent high temperature processing are expected to cause impurity segregation that could adversely affect the quality of Germanium films that are subsequently deposited on the Silicon. Therefore a process flow that can create highly doped source and drain pockets on a Silicon wafer that already has Ge layers on it, is preferable. The requirements on such a process flow are simple: the temperature during the fabrication should not approach 940 and in fact should be kept as low as possible to avoid the inter-diffusion of Ge and Si⁴⁸.

An ideal technique that can be adopted to achieve this goal is *Solid Phase Epitaxy (SPE)*^{49,50,51}. In SPE, the initial implantation of the Silicon with the desired dopant species is carried out with extremely high dosage ($\sim 1 \times 10^{15} \text{ cm}^{-2}$) so that the damage is high enough to completely *amorphize* the substrate. But the energy of the implant is kept low enough so as to restrict the amorphization to the superficial layers in the Silicon. If this is done successfully, then the Silicon underneath the damaged layer would still be single crystal. The importance of the single crystal Silicon is that it can act as a seed layer for the re-crystallization of the amorphous layers even at moderate temperatures, *thereby fixing the damage in the top Silicon and activating the dopants*. This recrystallization process is referred to as Solid Phase Epitaxy. It must be mentioned that the single crystal layers created by SPE are not completely devoid of point defects and those defects can increase the leakage current seen in transistors fabricated from the recrystallized substrates. The temperatures required for SPE are what make it extremely useful: the dopant activation and damage anneal can be accomplished at 500 to 700°C⁵² which is

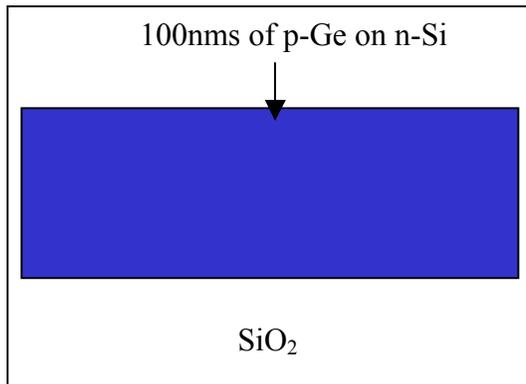
well below the thermal budget established for Germanium-on-Silicon. Therefore the use of SPE with a self-aligned fabrication process should produce Ge/Si Photo-HFETs with very little parasitic resistance. One problem is that since the implantation is done with low energy, the highly doped source and drain regions will obviously be superficial. But in the context of our devices that is not crucial since the channel is at the Si surface. The high surface doping will in fact aid in the formation of very low resistance source and drain contacts.

4.4.2 Fabrication

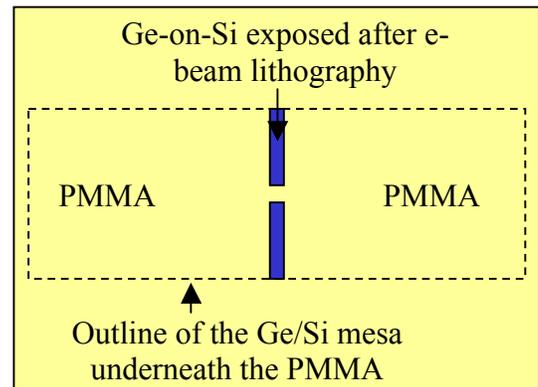
The process flow for the self-aligned devices incorporates solid phase epitaxy into the fabrication steps described in section 4.3.1 for the initial devices (the starting substrate again is a 100nm thick layer of p-Germanium on n-SOI). Fig 4-16 illustrates the various steps:

1. Like in section 4.3.1, the first step is to define a large Ge-on-Si mesa ($\sim 600\mu\text{m} \times 100\mu\text{m}$) using photo-lithography and dry etching. Germanium and Silicon are etched all the way to the buried oxide in an RIE chamber using CF_4 and a $\text{SF}_6/\text{C}_4\text{F}_8$ mixture respectively.
2. A thin layer of PMMA e-beam resist is spun on and patterned using electron-beam lithography to define the channel width and length. The channel is again made as a square and its dimensions are varied from $250\text{nm} \times 250\text{nm}$ to $4\mu\text{m} \times 4\mu\text{m}$.
3. Germanium and Silicon are dry etched all the way to the buried oxide using the same recipe as in step 1.
4. The residual PMMA is removed in acetone.

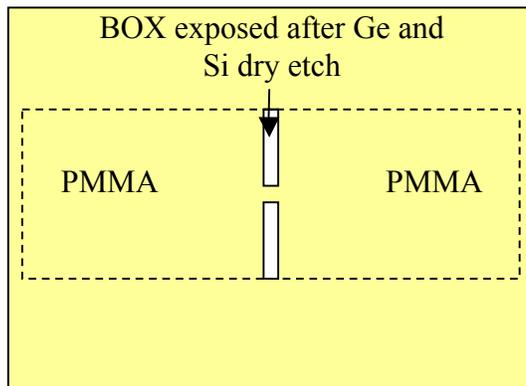
TOP VIEW (Steps 1 to 9)



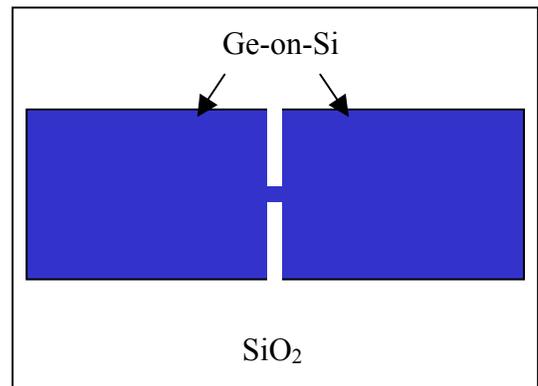
STEP 1



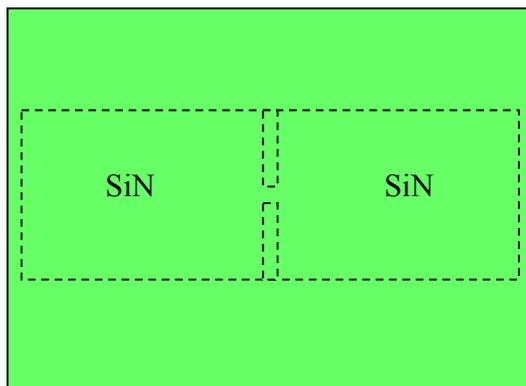
STEP 2



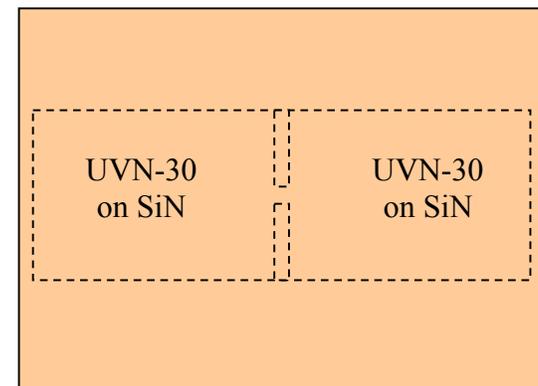
STEP 3



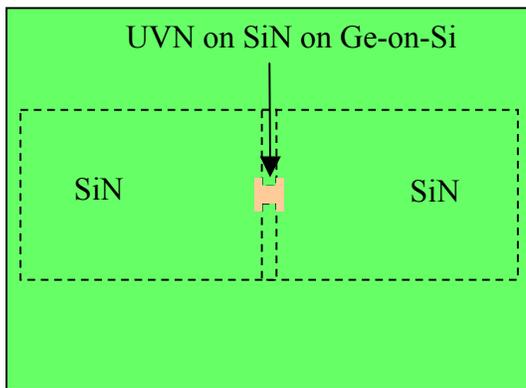
STEP 4



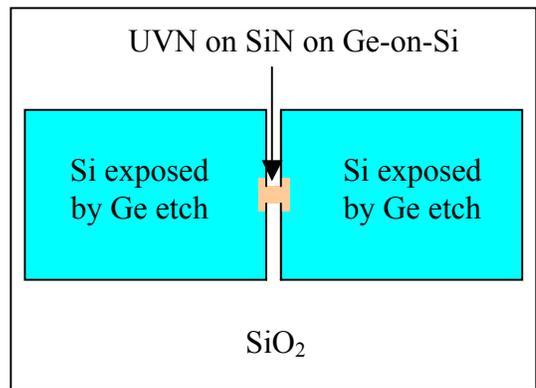
STEP 5



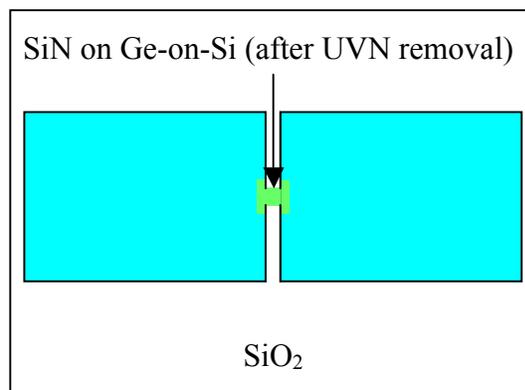
STEP 6



STEP 7

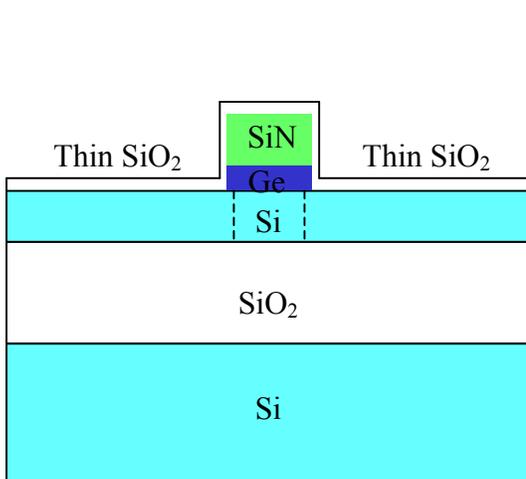


STEP 8

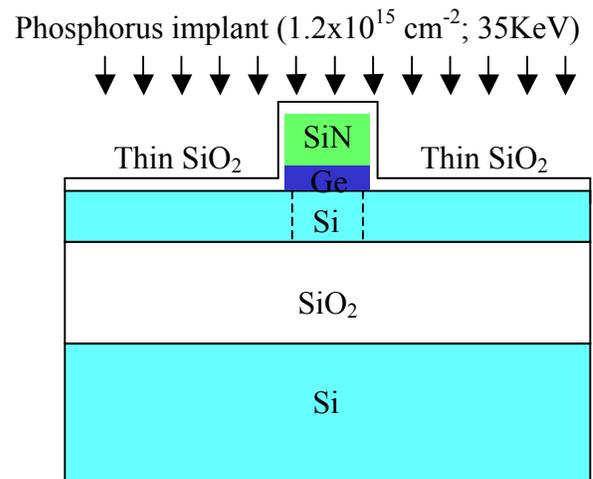


STEP 9

CROSS-SECTIONAL VIEW (Steps 10 to 14)



STEP 10



STEP 11(Part 1)

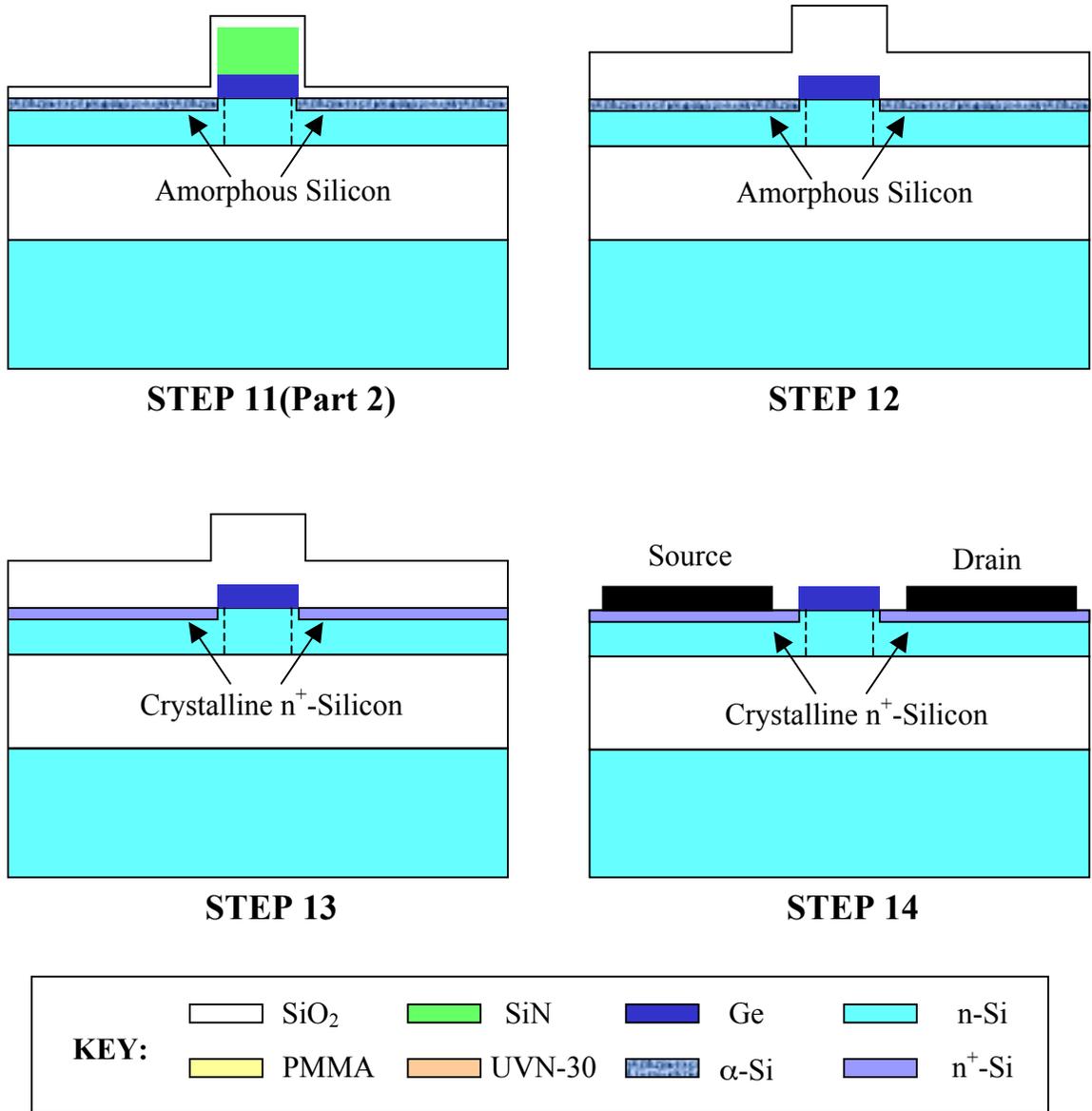


Fig 4-16: Fabrication steps involved in making self-aligned Photo-HFETs (For proper illustration, steps 1 to 9 are shown in top view and steps 10 to 14 are shown in cross-sectional view)

5. A 200nm thick film of Silicon Nitride is grown using a low temperature (300°C) PECVD process that leaves the Ge film unaffected. The nitride acts as the mask for the subsequent ion implantation step. Since the PECVD deposition produces an isotropic

layer of nitride, performing it after defining the channel *ensures that the sidewalls exposed by the channel etch are also protected by the nitride.*

6. A 400nm thick layer of UVN-30 (negative e-beam resist) is spun on the nitride.

7. The UVN is patterned by e-beam lithography, to define the Ge gate. The gate is aligned to the channel by using the same alignment marks during the exposure of the UVN and the PMMA. Since in a self-aligned process, the nitride covered Ge gate acts as the implantation mask, it is imperative that it completely cover the Si channel. If any part of the channel is exposed, it will get doped to the same level as the source and drain pockets thereby completely shorting it. Therefore, as opposed to the devices described in section 4.2, each side of the UVN gate-definition mask is made 200nm larger than the channel so as to increase the margin of error during the alignment. As illustrated by Fig 4-17, the 200nm increment increases the channel length by 400nm while leaving the width the same.

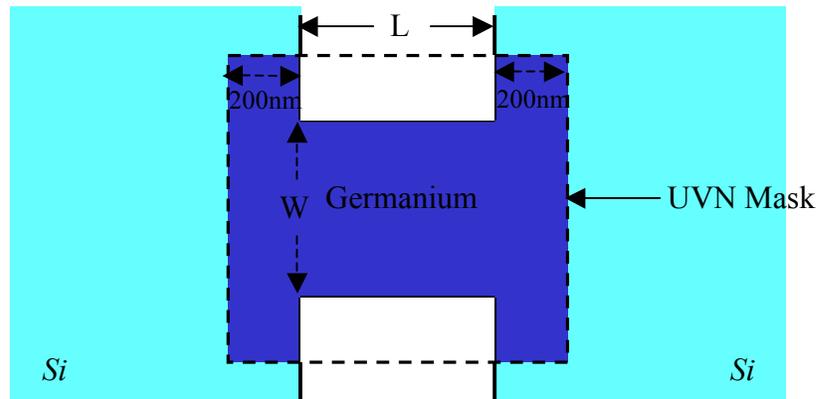


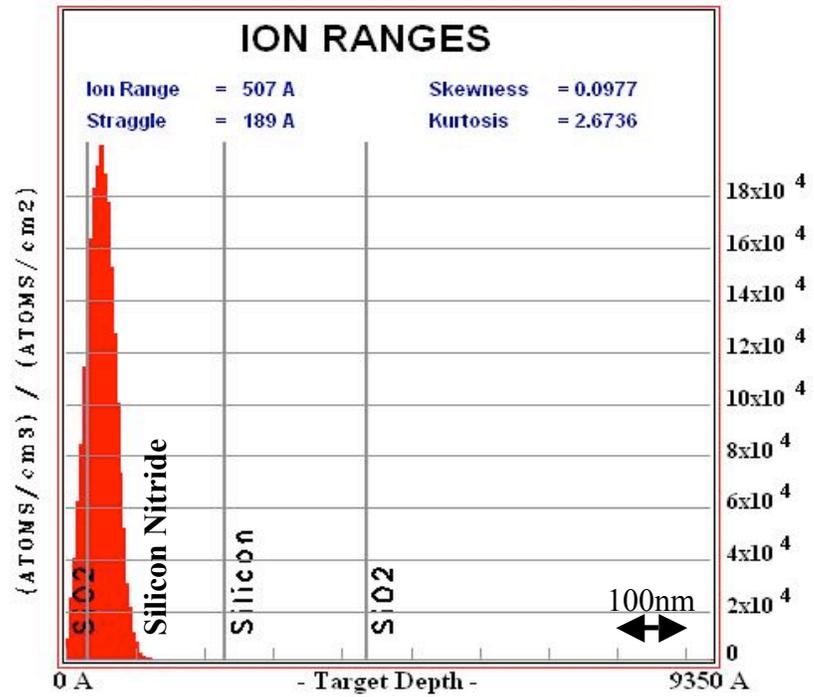
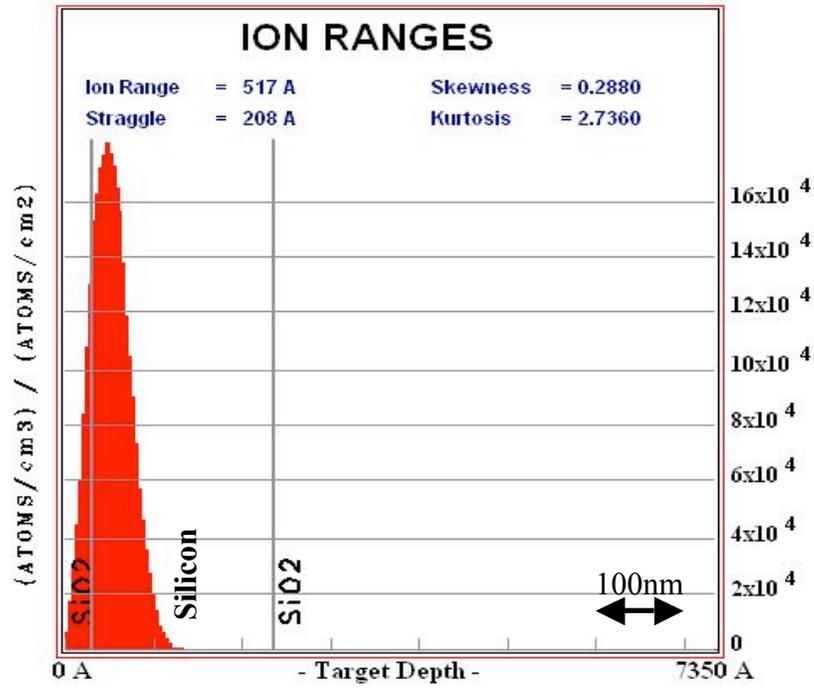
Fig 4-17: Top view of the gate. The UVN (dotted square) is patterned to be slightly bigger than the channel to facilitate alignment. After the etch we get the Ge gate shape shown in the figure.

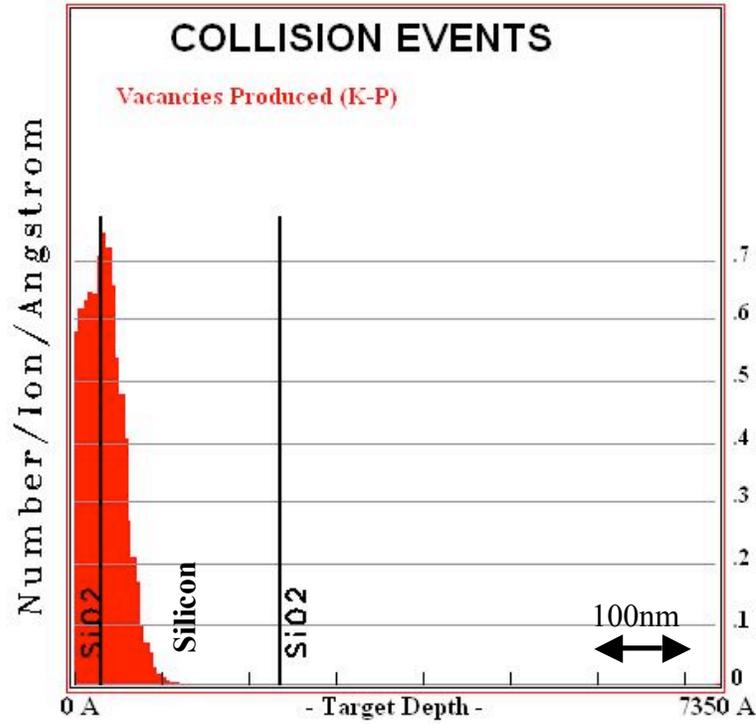
8. The gate is defined by dry etching the nitride and Germanium using CHF_3/Ar and CF_4 respectively. The last etch is carefully monitored so as to not harm the Silicon surface.

9. The residual UVN is removed in acetone.

10. A 30nm layer of PECVD oxide is deposited on the device. This acts to randomize the implanted ions in the next step thereby minimizing channeling effects. It also protects the Si surface from excess impurity adsorption.

11. Phosphorus ions are implanted ($1.2 \times 10^{15} \text{ cm}^{-2}$ dose; 35 KeV, 0° tilt) into the Silicon to define the source and drain. The expected range of penetration of the Phosphorus into the SiO_2 covered Silicon and the Silicon Nitride mask is shown in Figs 4-18(a) and (b). The calculations are done using standard SRIM software. Since the nitride is deposited using a low temperature PECVD process, it is expected to have significant Hydrogen contamination which can in turn lower its density. The target in the plot for Phosphorus implantation into nitride was therefore assumed to be SiNH (with a density of 2.2 g/cc) instead of Si_3N_4 . Since the maximum penetration depth of P^{31} in SiNH is found to be $< 100\text{nm}$, the 200nm nitride mask is clearly thick enough to ensure that none of the dopant atoms enter the Ge. Fig 4-18(c) shows that the peak density of vacancies produced in the Silicon is $\sim 0.7 /(\text{Incident Ion}).(\text{Angstrom})$. Therefore for a dose of $1.2 \times 10^{15} \text{ cm}^{-2}$, the density of vacancies would be $\sim 8 \times 10^{22} \text{ cm}^{-3}$, which far exceeds the amorphization threshold in the Silicon. Also from Fig 4-18(a) it is seen that the range of P^{31} ions in the Silicon is $\sim 50 \text{ nm}$, which means that the amorphization is indeed restricted to the superficial layers in the substrate.





(c)

Fig 4-18: Expected implantation profiles for Phosphorus (35KeV, no tilt) calculated using standard Stopping Range of Ions in Matter (SRIM) software. (a) Range of P⁺ ions in Silicon (b) Range of P⁺ ions in Silicon Nitride (c) Damage profile in Silicon in terms of number of vacancies produced.

12. After the implant, the 30nm of protective oxide and the residual Silicon Nitride layers are removed with a BOE etch and the sample is coated with 200nm of PECVD oxide, which acts as a protective layer during the anneal.

13. The sample is annealed at 650°C for 5 minutes in a Rapid Thermal Processing unit with a Nitrogen ambient. As mentioned in the introduction to SPE, even the relatively low-temperature anneal is expected to activate the dopants and cure the implant damage. Dummy SOI wafers that were also subjected to the implantation and anneal are used to

estimate the contact and sheet resistance of the actual substrate. A line of $100\mu\text{m} \times 100\mu\text{m}$ Ti/Al/Ti/Au ($25\text{\AA}/1500\text{\AA}/50\text{\AA}/1000\text{\AA}$) contacts separated by $200\mu\text{m}$ each are deposited on the dummy wafers (the Titanium is used for adhesive purposes; the Gold prevents the Al from oxidizing and aids in wire bonding) Resistance measurements between the various contacts show that the contact resistance is $\sim 100\Omega$ and the sheet resistance is $\sim 120 \Omega/\square$. These values indicate that the doping level in the source and drain is $\sim 1 \times 10^{20} /\text{cm}^{-3}$, which is expected from the dosage level used during the implant.

14. The protective oxide layer is removed and Ti/Al/Ti/Au contacts identical to the ones described in step 13 are deposited on the source and drain using photo-lithography and lift-off.

Fig 4-19 shows SEM micrographs of the finished devices. The samples are finally cleaved and wire-bonded onto high speed chip carriers that have a micro-strip active contact pad. The wire bonds are kept as small as possible in order to minimize the parasitic inductance in the circuit.

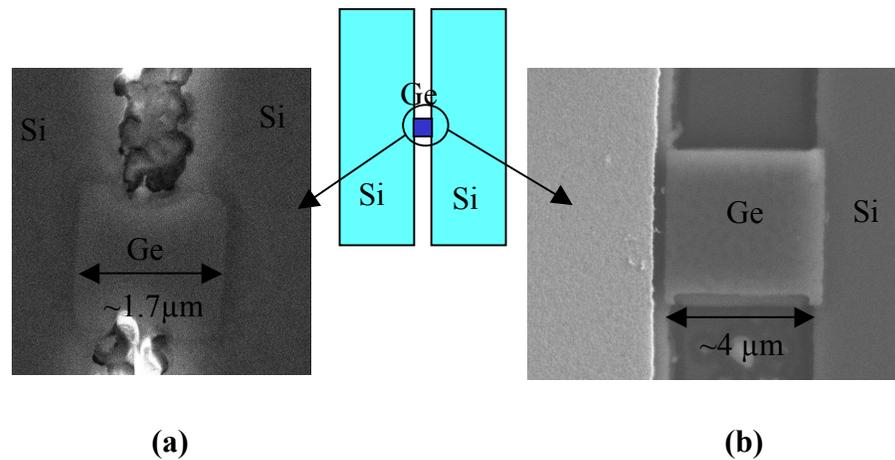


Fig 4-19: SEM images of a (a) $1.7\mu\text{m} \times 1.5\mu\text{m}$ and (b) $4.4\mu\text{m} \times 4\mu\text{m}$ device. In the $4\mu\text{m}$ device the bright layer on the left is the Gold contact.

4.5 Results: Pulsed response

4.5.1 Measurement Setup

The devices are characterized with normally incident light in a manner similar to the earlier experiments, but with a few important differences (Fig 4-20 shows a schematic of the setup). Since proper source and drain contacts have been defined, there is no need to use a 4-point measurement. Also, in order to carry out the high speed measurements the devices are connected through a bias-tee. The bias is applied through the inductive arm and the high speed signal is capacitively fed to a 20 GHz bandwidth sampling oscilloscope. The AC arm of the circuit is impedance matched to 50Ω throughout. This impedance is very crucial in determining the output voltage seen by the oscilloscope. This is because, as will be explained in the later sections, the expected change in channel current in these devices is only a few μA . Therefore the voltage seen at the 50Ω load is only of the order of a few $100\mu\text{V}$. Due to the low level of the signal, the SNR is improved by extensively averaging the output of the oscilloscope using an externally connected computer ($\sim 100,000$ waveforms are used in each average). The noise is also reduced by employing a Faraday cage around the sample (and the focusing optics) in order to shield it from extraneous parasitic sources.

In the optical part of the setup, a microscope objective is used to focus $1.55\ \mu\text{m}$ light onto the Germanium. The beam waist at the focal point is $\sim 10\mu\text{m}$. An IR camera captures the reflection from the device and the image is used to better align the laser to the gate.

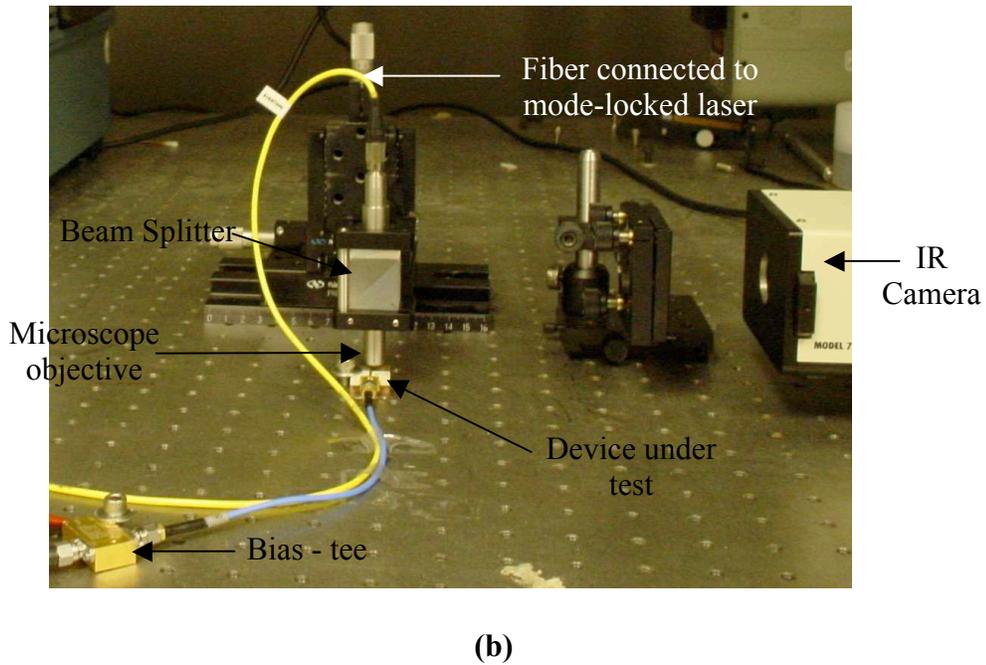
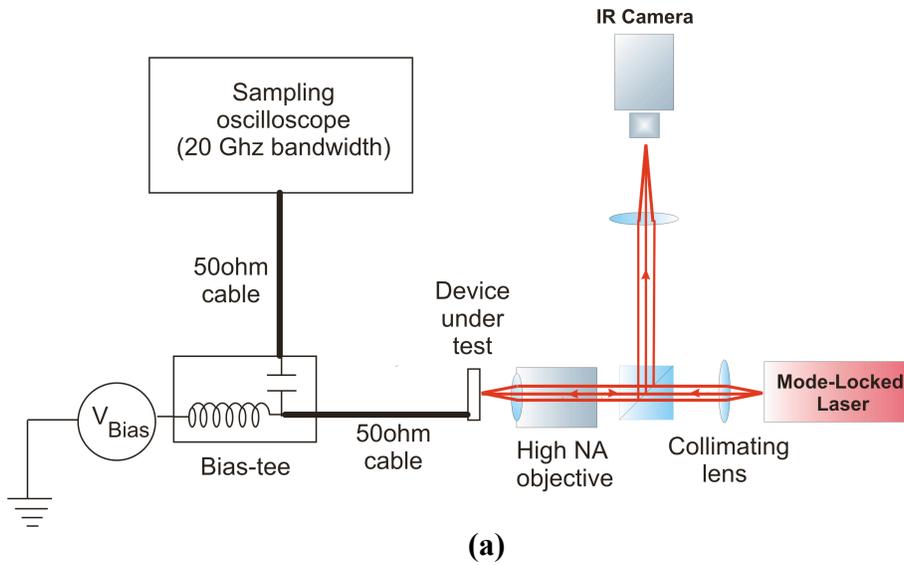


Fig 4-20: (a) Schematic of the high speed measurement setup (b) Photograph of the actual setup.

4.5.2 DC Electrical Characterization

Fig 4-21 shows the I_{DS} - V_{DS} curves for two devices with different dimensions: channel length and width in one are $1.7\mu\text{m}$ and $1.5\mu\text{m}$ and in the other, $1.15\mu\text{m}$ and

0.5 μm (the processing in the devices was done to obtain dimensions of 1.4 μm x 1 μm and 0.9 μm x 0.5 μm respectively, but inconsistencies in the e-beam lithography resulted in the actual dimensions listed above). Even though typical FET channel current saturation is seen, it should be noted from the plots that the 1.7 μm x1.5 μm device is leakier than anticipated (the reason for the high leakage current and its comparison to the theoretical value is discussed in the analysis in section 4.6.1). But in spite of the higher leakage, due to overall fabrication yield, most of the pulsed response measurements described in the next section are performed on this device. It is also worth mentioning that the subsequent high speed experiments are all carried out at 0.5V bias, which provides a good compromise between dark current and photo-response.

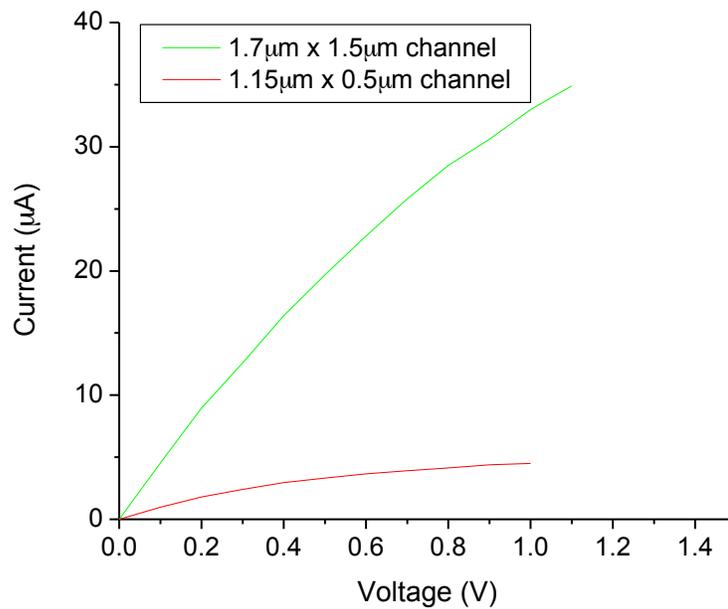
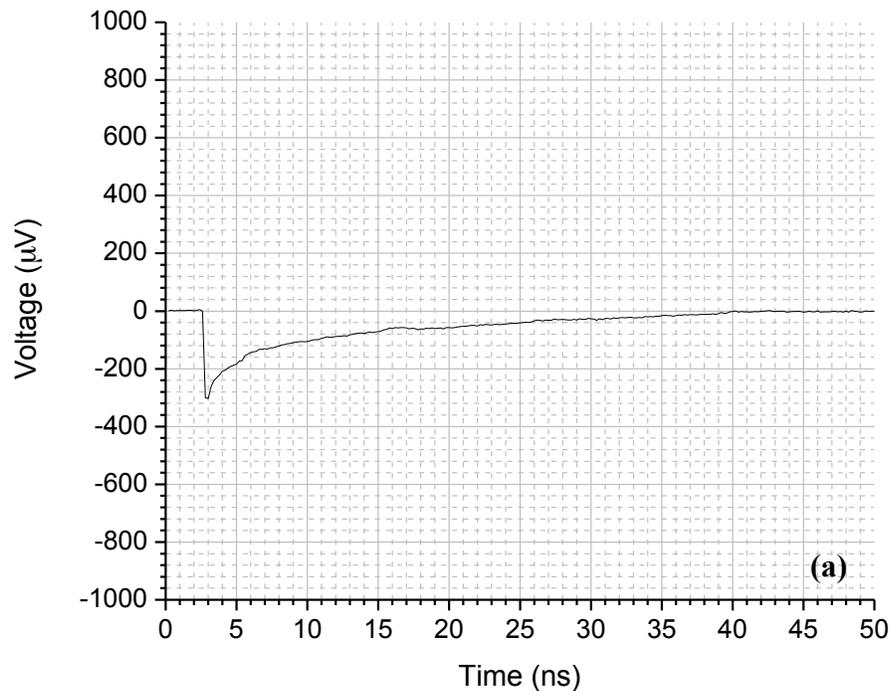


Fig: 4-21: Channel current vs Source-Drain bias voltage for Photo-HFETs of two different sizes (1.7 μm x1.5 μm and 1.15 μm x0.5 μm). Even though current saturation is seen, it is less than expected in the larger device.

4.5.3 Pulsed Response

Fig 4-22 shows the response of the $1.7\mu\text{m} \times 1.5\mu\text{m}$ device to pulsed light. The source is a mode-locked laser that produces 1ps pulses with a peak power of $\sim 50\text{W}$. The repetition rate of the laser is 20 MHz and it therefore has an average power of $\sim 1\text{mW}$. Fig 4-22(a) plots the response over 50ns which is the time between pulses. *It can be clearly seen that the rising edge in the device is extremely sharp and the fall time is very long.* Fig 4-22(b) shows a higher resolution picture of the rising slope; it can be seen from the plot that the rise time is $\sim 40\text{ps}$. (The rise time represents a fall in voltage because of the way the bias circuit is set up. When light is incident on the gate the resistance of the channel decreases and a spike of current goes through it. But this current goes through the 50Ω load from ground to positive, thereby causing a voltage drop to be recorded by the oscilloscope).



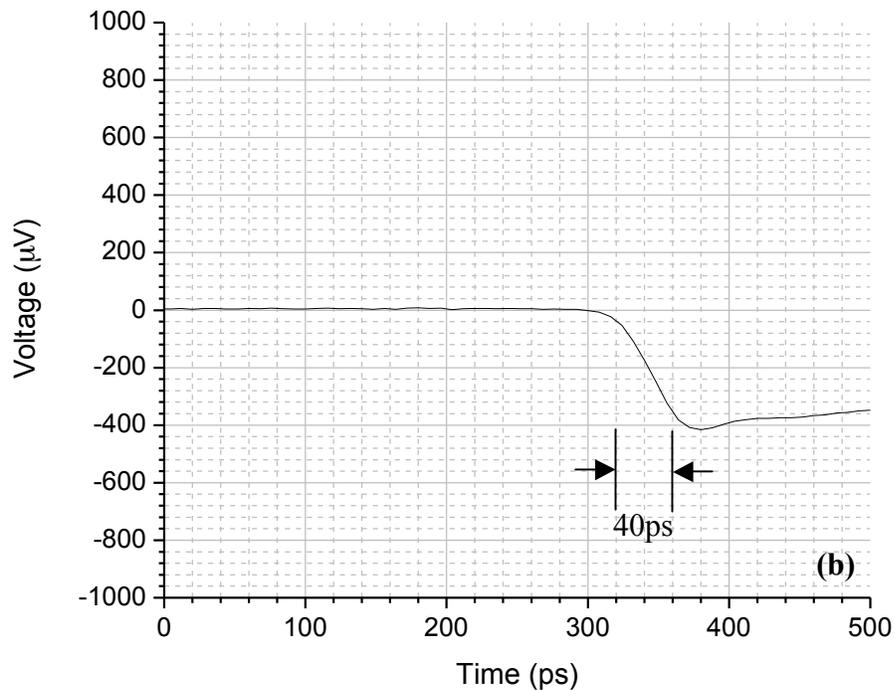


Fig 4-22: Response of the 1.7µm x 1.5µm device to 1ps mode-locked pulses @1.55 µm (~1mW average power; 20MHz rep rate), recorded by the 50Ω input of a 20GHz sampling oscilloscope. The data was averaged >100,000 times to reduce noise. (a) Response over the whole 50ns duration between two pulses (b) Zoomed in picture of the rising edge.

These time domain features of the Photo-HFET are extremely important and are the main conclusions of the high speed measurements. The sharp rise time is especially relevant as it indicates that this device design can be successfully integrated into high bandwidth transceivers operating at > 10Gbps. The slow fall time is a significant caveat, but it can be tackled by improved device structures and adaptive circuit designs. The analysis in the next section takes an in-depth look at these and other features of the devices.

4.6 Analysis of detector response

A crucial aspect of the device that must be looked at thoroughly is the expected photo-response. In order to perform the analysis successfully, an accurate model for the transistor channel current is essential. Even though commercial simulators (like the ATLAS package mentioned below) can ably predict the behavior of conventional FETs, they are not ideally suited to analyzing the effect of light on an unconventional device design like the one being discussed here. Therefore a physically precise model for FET structures is required that can be adapted to include the effect of the incident optical pulses so as to estimate the expected change in channel current due to the light.

4.6.1 Channel current saturation in JFETs

As mentioned previously, in the absence of light, the detector behaves like a normal JFET. The channel current starts displaying some saturation at a source-drain bias $>0.3\text{V}$ (as shown in Fig 4-21). The saturation occurs because a positive voltage on the Silicon channel reverse biases the p-Ge/n-Si junction thereby increasing the depletion in the Silicon and pinching off the channel. Therefore when a bias is applied across the source and drain, the channel starts pinching off at the higher voltage side.

The current saturation effect is illustrated in Fig 4-23, which shows the expected I_D - V_D plot extracted from SILVACO's standard semiconductor device simulator ATLAS®, super-imposed over the experimental values. The calculation is done for both, the $1.7\mu\text{m} \times 1.5\mu\text{m}$ and the $1.15\mu\text{m} \times 0.5\mu\text{m}$ devices (dimensions are channel length x width). In the simulation, the Germanium is assumed to have a p doping of $1 \times 10^{18} \text{ cm}^{-3}$,

the Silicon channel doping profile is taken from the SIMS results plotted in Fig 4-6 and the source and drain pockets are assumed to be doped to $5 \times 10^{19} \text{ cm}^{-3}$.

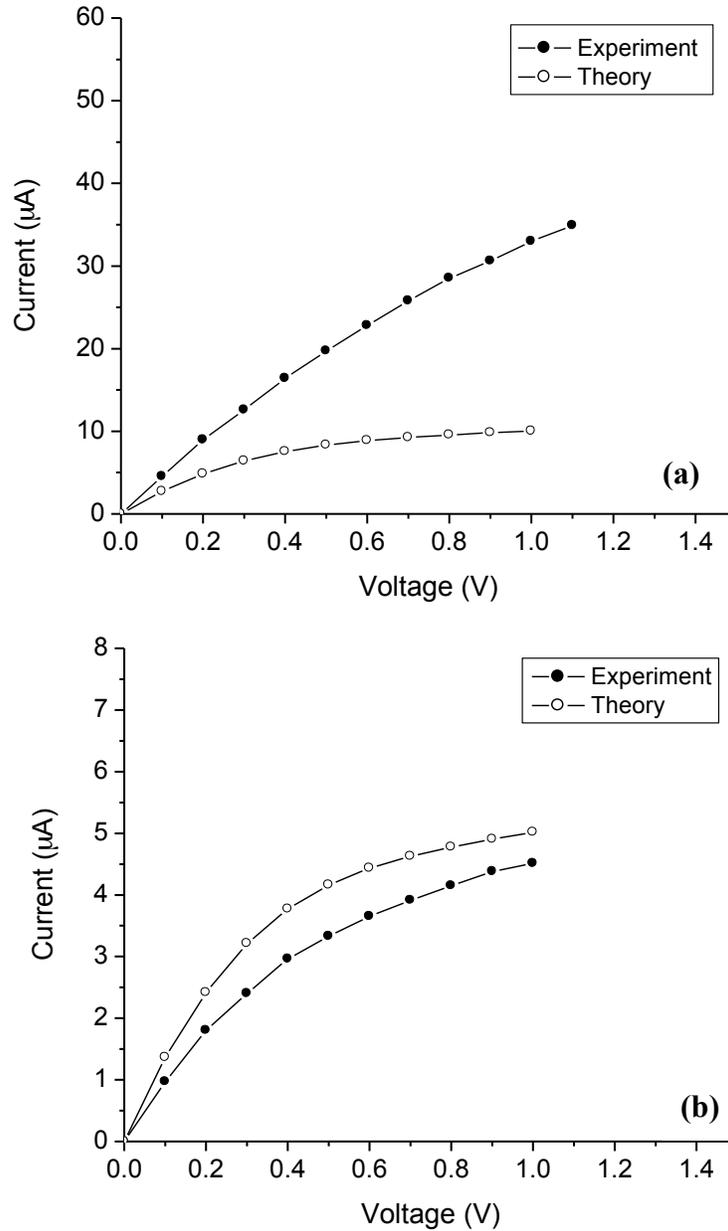


Fig 4-23: Experimental channel current vs source-drain voltage curves compared with the expected plots (simulated using ATLAS) (a) 1.7µm x 1.5µm channel length x width (b) 1.15µm x 0.5µm channel length x width. The larger device shows higher leakage due to fabrication imperfections.

Here it can be clearly seen that the expected amount of current saturation is not seen in the larger device (Fig 4-23(a)). The higher channel current can be attributed to improper channel definition during steps 2 and 3 of the process flow described in section 4.4.2. The exposure during the e-beam lithography was inconsistent due to which certain areas were left unexposed, thereby creating narrow shorting paths across the channel at a few places (Fig 4-24).

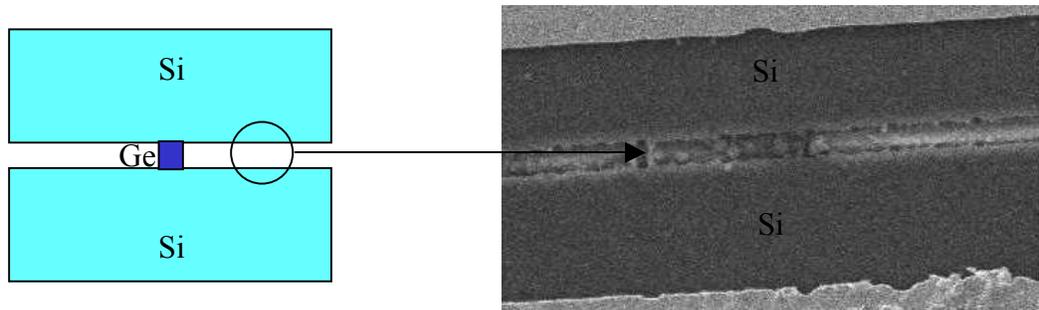


Fig 4-24: Top view schematic of the device and SEM picture of the etched gap in the Silicon that defines the channel (and ideally limits its conductance path to the section under the Ge gate). Inconsistencies in the e-beam lithography created narrow shorts across the channel as pointed to by the arrow.

4.6.2 Modeling JFET channel current without light

The standard model for analyzing the $I_{DS}-V_{DS}$ curves in JFETs is very well understood^{29,44}. As described above, an increase in the source-drain bias increases the channel depletion close to the electrode that is at the higher potential. Fig 4-25(a) shows a schematic picture of the expected depletion profile.

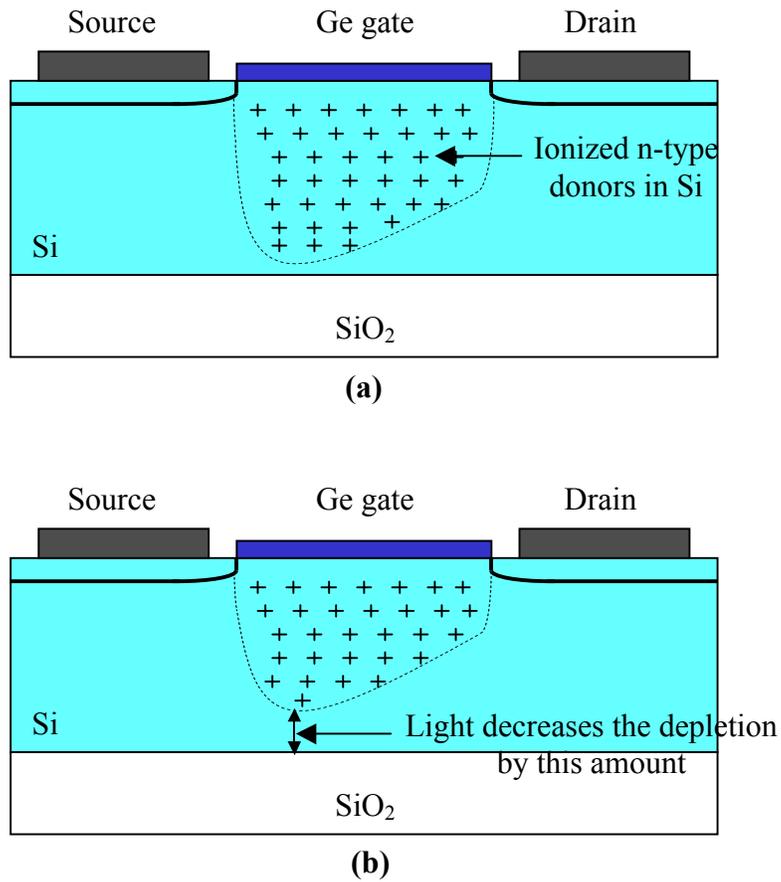


Fig 4-25: Schematic diagram of the depletion profile in the Si channel (a) without and (b) with light. The source is assumed to be at a higher voltage due to the source-drain bias (the drain is grounded).

The change in depletion caused by the mode-locked laser in our devices is ~20 to 30nm.

Assuming an abrupt one-sided junction between the gate and the channel (which accurately describes the p^+ -Ge/n-Si junction in our device), the depletion width at a point x along the channel is given by the standard expression:

$$h(x) = \left(\frac{2 \epsilon_s [V(x) - V_G + V_{bi}]}{q N_D} \right)^{1/2} \quad (4.5)$$

where, N_D is the doping in the Si channel, $V(x)$ is the source bias voltage at x with respect to the drain, V_G is the gate voltage and V_{bi} is the built-in voltage at the junction ($\sim 400\text{mV}$). In the Photo-HFET the gate voltage is the forward voltage applied by the incident light. It must be mentioned here that due to its high doping, the Germanium is always assumed to be an equi-potential film that floats to the appropriate potential determined by the overall charge distribution.

Now, the channel current density at a point x in the channel is given by

$$J_x = qN_D \mu_n E_x \quad (4.6)$$

where, E_x is the local electric field in the x -direction and μ_n is the mobility of electrons in the channel. μ_n can be assumed to be field independent at the bias levels used in the experiments. Substituting $E_x = -dV/dx$ and converting to absolute current by multiplying with the area of the conductance path at x , we get

$$I_{Chan} = qN_D \mu_n \left(\frac{dV}{dx} \right) [a - h] W \quad (4.7)$$

where $[a-h] W$ is the conductive area (a is the total thickness of the Si channel and W is the width of the device). Therefore,

$$I_{Chan} dx = W q N_D \mu_n [a - h] dV \quad (4.8)$$

Now to integrate eqn 4.8, dV has to be replaced by a function of h . This can be done by differentiating eqn (4.5):

$$dV = \frac{qN_D}{\epsilon_s} h dh \quad (4.9)$$

Substituting this expression into eqn 4.8 and integrating, we obtain:

$$I_{Chan} = \left(\frac{1}{L}\right) \int_{h_D}^{h_S} W q N_D \mu_n (a - h) \left(\frac{q N_D}{\epsilon_S}\right) h dh \quad (4.10)$$

where, the limits of the integration h_D and h_S represent the depletion widths at the drain and source ends of the channel respectively. h_S is obtained by putting $V(x) = V_{Bias}$ in eqn 4.5 and similarly h_D is obtained by putting $V(x) = 0$. Evaluating the integral in eqn 4.10, we finally get the channel current in terms of the depletion width at the two ends:

$$I_{Chan} = \frac{W \mu_n q^2 N_D^2 a^3}{6 \epsilon_S L} \left[\frac{3}{a^2} (h_S^2 - h_D^2) - \frac{2}{a^3} (h_S^3 - h_D^3) \right] \quad (4.11)$$

Before applying this equation to the fabricated devices and comparing the output to the experimental data, an important clarification must be made. In the above derivation, the channel is assumed to have uniform doping (N_D) throughout. Unfortunately, as shown in the SIMS profile in Fig 4-6, the SOI substrates used in the fabrication have a small peak in doping at the surface (of $\sim 1 \times 10^{17} \text{ cm}^{-3}$) which then falls off to $\sim 2.5 \times 10^{16} \text{ cm}^{-3}$ in the bulk. Due to this, the expression for $h(x)$ (eqn 4.5) is not valid exactly. In fact the relationship between the voltage across the Ge/Si junction and the depletion in the Silicon takes on a rather more complicated form that cannot be integrated analytically. However, as the higher doped region close to the surface is very

thin ($\sim 20\text{nm}$), the above formulation should still give a reasonably accurate estimate of the channel current if we assume a uniform doping of $2.5 \times 10^{16} \text{ cm}^{-3}$ in the Silicon.

An additional minor adjustment can also be made to further improve the accuracy of the model: Since the higher surface doping serves to slightly reduce the depletion width in the Silicon and therefore increase the thickness of the undepleted conductive path in the channel, its effect on the current can be taken into account by assuming a slightly higher ‘equivalent’ channel thickness for the device. For example, in the 205 nm thick substrates used in the fabrication (with a $1 \times 10^{17} \text{ cm}^{-3}$ doping spike at the surface), a bias of 0.5V gives a source depletion width, h_S , of $\sim 190 \text{ nm}$. Therefore there is a $\sim 15\text{nm}$ thick undepleted current path at the source. Now by comparison, a channel with a uniform doping of $2.5 \times 10^{16} \text{ cm}^{-3}$ will have a source depletion width of $\sim 210 \text{ nm}$ for the same biasing conditions (this is obtained by setting $V(x)=0.5$ and $V_G=0$ in eqn 4.5). Therefore for this case an ‘equivalent’ channel width of $\sim 225\text{nm}$ would give the same 15nm thickness for the undepleted current path. In other words a 225nm thick ‘equivalent’ channel that is uniformly doped at $2.5 \times 10^{16} \text{ cm}^{-3}$ will have roughly the same conductance as a 205nm thick channel with the same bulk doping but with a small peak of $1 \times 10^{17} \text{ cm}^{-3}$ doping at the surface.

Keeping this in mind, the values predicted by the dark current model can now be compared to the ones obtained from the ATLAS simulation and the experiments. For $N_D = 2.5 \times 10^{16} \text{ cm}^{-3}$ and $V_{\text{Bias}} = 0.5\text{V}$, eqn 4.5 gives $h_S \sim 210\text{nm}$ and $h_D \sim 140\text{nm}$. Plugging these into eqn 4.11 along with $a=225\text{nm}$, $W = 1.5\mu\text{m}$, $L = 1.7\mu\text{m}$ and $\mu_n = 1000 \text{ cm}^2/\text{Vs}$, gives $I_{\text{Chan}} = 8.25\mu\text{A}$. *This is extremely close to the value predicted by the commercial*

software ($8.3\mu A$) [as explained in section 4.6.1, the experimental value is a little higher ($\sim 11.9\mu A$) due to fabrication problems]. Since this establishes the validity of the model, we can now proceed to calculating the light induced change in current.

4.6.3 Magnitude of the photo-response: Estimating change in channel current

The expected change in channel current can be modeled by calculating the gate voltage, V_G , created by the light incident on the device. A rough estimate for V_G can be arrived at by using $V_G \approx Q_{\text{Optical}}/C_{\text{dep}}$ where C_{dep} is the gate depletion capacitance and Q_{Optical} is the total amount of charge induced on the gate by the mode-locked pulses. Even though this expression is very intuitive, it has its limitations due to the fact that it doesn't account for the change in capacitance that accompanies the change in depletion caused by the light. Therefore we take a slightly more rigorous approach towards estimating V_G by *first calculating the change in depletion and then converting that to a gate voltage*.

Unlike the CW regime that was discussed in section 4.3.3, the change in channel depletion caused by a 1ps pulse cannot be analyzed by extracting the induced forward-bias from the junction I-V curve. This is because the time scale of the pulse is too short for the I-V curve to accurately describe carrier transport across the p-Ge/n-Si junction. Hence we use a more straightforward method of evaluating the change in depletion width: it is calculated from the *absolute number* of photo-electrons that successfully diffuse into the Silicon from the Germanium. This is because the 'collected' photo-electrons can be thought to neutralize the ionized donors in the n-Si, thereby reducing the amount of depletion. This way of looking at things is not very useful in the CW case because there only the *rate* of photo-electron generation and collection is known and it is

therefore much more convenient to analyze the device in terms of currents instead of absolute number of charge carriers.

Equating the number of collected photo-electrons to the number of neutralized donors in a region that has a thickness of Δh , where Δh represents the change in depletion, we get:

$$\Delta h = \frac{\phi}{N_D A_{Gate}} \quad (4.12)$$

Here, ϕ is the total number of collected photo-electrons, N_D again is the doping in the Silicon ($2.5 \times 10^{16} \text{ cm}^{-3}$) and A_{Gate} is the area of the Ge gate ($1.5 \times 1.7 \approx 2.5 \mu\text{m}^2$). Similar to eqn 4.3, ϕ is simply the number of photo-carriers created within one diffusion length of the interface. Therefore,

$$\phi = \frac{P_{Avg} \tau_{rep} (1 - R_{ref}) (1 - e^{-\alpha L_D})}{(h c / \lambda)} \frac{A_{Gate}}{A_{Spot}} \quad (4.13)$$

where P_{Avg} is the average power in the mode-locked laser, τ_{rep} is the duration between pulses and the other symbols have their usual definition.

The magnitude of the forward bias induced on the gate, V_G , is calculated from Δh by again invoking the depletion width expression for a one-sided junction (eqn 4.5):

$$\sqrt{(V(x) + V_{bi})} - \sqrt{(V(x) + V_{bi} - V_G)} = \Delta h \sqrt{\frac{q N_D}{2 \epsilon_S}} \quad (4.14)$$

where, the definitions of all the variables are the same as in eqn 4.5.

Here, it must be stressed that since the Germanium is an equi-potential film, V_G is independent of the length coordinate x . Also, since the gate voltage represents a forward

bias caused by the photo-electrons neutralizing ionized donors in the Silicon, its effect is to simply lower the potential of the Si channel by V_G *throughout*. Therefore the difference between the source and drain potential always stays at V_{Bias} irrespective of how much gate voltage is applied.

Since V_G is constant throughout the length of the channel, it can be seen from eqn 4.14 that ideally Δh varies in accordance with $V(x)$ in order to keep V_G independent of x . A key element in the variation of Δh with $V(x)$ is that it will be at its highest for $V(x) = 0$. *Therefore if a rough estimate is known for the maximum possible value of Δh , then an approximate value for V_G can be calculated from that by putting $V(x) = 0$ in eqn 4.14.*

In order to find a maxima for Δh we first need to estimate an upper bound on the number of collected photo-electrons, ϕ , from eqn 4.13. In our experiment P_{avg} is $\sim 0.8\text{mW}$ (the output of the laser is 1mW and there is $\sim 1\text{dB}$ of loss in the optics), $\tau_{\text{rep}} = 50\text{nS}$, $R_{\text{Ref}} = 0.4$, $A_{\text{Gate}} = 2.5\mu\text{m}^2$, and $A_{\text{Spot}} = 80\mu\text{m}^2$. The value of the quantum efficiency, $1 - \exp[-\alpha L_D]$ is ~ 3 to 4×10^{-4} in (from section 4.2.2). Plugging these values into eqn 4.13, we get $\phi \approx 2000$ carriers. Using this value in eqn 4.12, Δh is found to be $\sim 30\text{nm}$. Finally, substituting this and $V(x) = 0$ in eqn 4.14 gives $V_G = 0.15\text{V}$.

The above formulation has established that the incident mode-locked pulses forward bias the gate-channel junction by raising the potential of the Si by $\sim 0.15\text{V}$. This gate voltage can be converted to a change in channel current by updating the values of the depletion widths h_S and h_D in eqn 4.11, to account for the 0.15V forward bias. The new values of h_S and h_D can be extracted from eqn 4.5: Putting $V_G = 0.15\text{V}$, $V_{\text{bi}} = 0.4\text{V}$ and $V(x) = 0\text{V}$ gives $h_D \sim 110\text{nm}$. Similarly putting $V_G = 0.15\text{V}$, $V_{\text{bi}} = 0.4\text{V}$ and $V(x) = 0.5\text{V}$

gives $h_S \sim 190\text{nm}$. Fig 4-25(b) schematically depicts the opening of the channel caused by the forward bias.

Plugging the above values into the expression for channel current (eqn 4.11), we see that in the presence of the light pulses, $I_{\text{Chan}} \sim 12.1 \mu\text{A}$ for a source-drain bias of 0.5 V. Since the dark current is $8.2\mu\text{A}$, the expected change in current measured at the output is $\sim 4\mu\text{A}$. It can be seen that this agrees reasonably well with the experimental data shown in Fig 4-22, where the measured change in current is $\sim 8 \mu\text{A}$ ($400\mu\text{V}$ across a 50Ω load)..

In order to make the comparison between the theory and experiments at an additional data point, the results for a smaller device ($1.15\mu\text{m} \times 0.5\mu\text{m}$) are also analyzed. Following the exact same procedure used above, the expected change in channel current for this device is found to be $\sim 1.95\mu\text{A}$. The measured value is $\sim 1.2\mu\text{A}$ and hence we again see fairly good agreement between the two. (The yield of the $1.15\mu\text{m} \times 0.5\mu\text{m}$ devices was such that the pulsed experiments could not be performed with sufficiently short wire bonds. Hence the magnitude of the response could be measured reasonably accurately for these devices, but the rise time could not. Instead of being intrinsically limited, the rise time gets restricted to $\sim L/50\Omega$, where L is the parasitic inductance of the wire bond. Further details about the various time constants in the device will be discussed later in this chapter).

A final observation that is worth making is that since the induced forward bias (0.15V) is much less than the built-in voltage, the response of the detectors is expected to vary almost linearly as the optical power is lowered. This is verified by attenuating the

incident power on the larger device in steps of 3dB all the way to 18dB and the change in channel current is found to be roughly linear throughout.

Therefore to conclude, we now have a good theoretical framework based on normal JFET behavior that predicts the pulsed response of the fabricated Photo-HFETs with a reasonable degree of accuracy.

5.6.4 Analysis of the rise time

It can be seen from Fig 4-22(b), that the rise time in the 1.7 μ m detector is \sim 40ps. This rise time is limited by two factors: the time it takes for the generated photo-carriers to diffuse from the Germanium into the Silicon and the time needed for the external circuit to fully detect the increased channel conductance made available by the light. The diffusion time of the electrons is essentially their lifetime in the Germanium, since only the electrons within one diffusion length of the Ge/Si interface get collected. Since L_D is extremely small (\sim 10nm), the lifetime ($\tau_n = L_D^2/D_n$) is very short. Even if the mobility in the heavily dislocated Germanium is about 100cm²/Vs, τ_n is still only of the order of 1 ps. Therefore the second component of the rise time described above is the one that dominates the response. In accordance with normal FET behavior, that rise time is simply determined by the transit time for electrons across the channel. In order to make a rough estimate for the transit time (τ_T) we can assume a uniform electric field throughout the device, $E_x = V_{Bias}/L$. Hence τ_T is given by :

$$\tau_T = \frac{L}{\mu E_x} \approx \frac{L^2}{\mu V_{Bias}} \quad (4.15)$$

For a 1.7 μm device the expected transit time is of the order of 50ps which is what was seen experimentally.

The short rise time in these devices is very suitable for use in high speed transceivers operating at $\sim 10\text{Gbps}$. As with any transistor structure, the speed will improve as the device is made smaller. Since the diffusion time of the carriers does not put any constraint on the bandwidth, devices that are in the 100nm channel length range should have rise times of only about 1ps ($100\text{nm}/v_{\text{sat}}$, where the saturation velocity v_{sat} is $\sim 1 \times 10^7$ cm/s). As conventional MOSFETs have already been scaled to smaller dimensions than 100nm it is clear that the Photo-HFET design can be made fast enough to be compatible with higher bandwidth transceivers well into the future.

In the context of the bandwidth it is also worth noting that the cut-off frequency of the devices *is simply the reciprocal of the rise time*. This can be seen as follows: for a JFET structure, the cutoff frequency is given by:

$$f_T = \frac{g_m}{2\pi C_G} \quad (4.16)$$

where, $g_m = \partial I_{\text{Chan}} / \partial V_G$ is the transconductance and C_G is the gate capacitance. This is the frequency at which the current in the gate circuit becomes equal to the channel current (i.e., current gain becomes unity). Now g_m can be rewritten as:

$$g_m = \frac{\partial I_{\text{Chan}}}{\partial V_G} = \frac{\partial I_{\text{Chan}}}{\partial Q_G} \frac{\partial Q_G}{\partial V_G} = \frac{\partial I_{\text{Chan}}}{\partial Q_{\text{Chan}}} \frac{\partial Q_G}{\partial V_G} \quad (4.17)$$

where, the last equality occurs because a *change in the charge across the depletion capacitance directly translates into a change in charge available to the channel*. Now,

$\partial Q_G/\partial V_G$ is C_G by definition and $\partial Q_{\text{Chan}}/\partial I_{\text{Chan}}$ is the transit time because it represents the characteristic time for charge transport across the channel. Numerically this relationship can be arrived at easily: we know that $I_{\text{Chan}} = qN_D v A_{\text{Cross}}$ where v is the electron velocity and A_{Cross} is the cross-sectional area of the channel. Multiplying and dividing by channel length L gives us $I_{\text{Chan}} = (qN_D A_{\text{Cross}} L) (v/L) = Q_{\text{Chan}} \tau_T$.

Therefore plugging the values of the two derivatives into eqn 4.17 we get:

$$g_m = \frac{C_G}{\tau_T} \Rightarrow \frac{g_m}{C_G} = \frac{1}{\tau_T}$$

Comparing with eqn 4.16, we get the desired result: $f_T = 2\pi/\tau_T$ or $\omega_T = 1/\tau_T$.

In a JFET, g_m can be estimated from the channel current equation (eqn 4.11) by expressing the depletion widths h_S and h_D in terms of V_G and then differentiating. Doing that we obtain:

$$g_m = \frac{2W \mu_n q N_D (h_S - h_D)}{L} \quad (4.18)$$

For the values of h_S and h_D found in section 4.6.3, we get $g_m \approx 5 \times 10^{-5}$ Siemens. C_G can be estimated simply as $\epsilon_S A_{\text{Gate}}/h$, where $A_{\text{Gate}} = 2.5 \mu\text{m}^2$ and the depletion width h is of the order of 100nm. Therefore C_G for the fabricated device is ≈ 2.5 fF. Plugging these values into eqn 4.16 gives $f_T \sim 3.18$ GHz, or in other words the characteristic time constant C_G/g_m is ~ 50 ps. This compares quite well to the transit time calculated earlier in the section.

Now, if the measurement time scale is $\ll 50$ ps, the frequency would be well beyond f_T and so the gate current will be much larger than the induced change in channel current. In the Photo-HFET, the “gate current” is essentially the photo-current created by

the collection of the photo-electrons. Since the carrier-collection period is the exceedingly short lifetime of the electrons in the Ge, the time scale for the flow of the gate current is determined by the duration of the incident pulse. Therefore $I_G \approx q\phi/\tau_{\text{pulse}}$, where I_G is the gate current and ϕ is again the number of collected photo-electrons.

In our experiments the pulse width is only 1ps and therefore we do find that $I_G \gg \Delta I_{\text{Chan}}$ ($I_G \approx 320\mu\text{A}$, $\Delta I_{\text{Chan}} \approx 8\mu\text{A}$). *Hence the current gain in the measured device is much less than 1.*

5.6.5 Analysis of the fall time

The fall time, τ_F , of the photo-response is the average recombination time of the trapped photo-holes at the Ge/Si interface (therefore, in essence it represents the ‘dielectric relaxation’ time in the Germanium). The appropriate interpretation of τ_F is that it is the RC discharging time for the depletion capacitance at the p-n junction. This is because the charge-separation created by the diffusion of the photo-electrons charges the depletion capacitance which then has to discharge through the forward resistance of the diode. The discharging process itself can be understood as follows: as has been discussed in great detail previously, the light generated charge-separation creates a small forward bias across the open p-Ge/n-Si junction. This results in the flow of a forward current that provides electrons to neutralize the trapped photo-holes at the Ge/Si interface.

It is very important to realize that τ_F *is not the same as the minority carrier recombination time* in the Germanium, which is exceedingly low ($\sim 1\text{ps}$). By contrast, the

dielectric relaxation time is much longer because there is initially a lack of electrons with which the excess holes can recombine.

Since the resistance through which the depletion capacitance discharges is the forward resistance of the diode, it can be written as $R_F \approx \eta V_{th} / I_F$, where I_F is the forward current. This value of R_F comes directly from the diode equation (eqn 4.1) after making the assumption that the forward current is much greater than the dark current in the junction. In section 4.6.3 we estimated that the forward bias created by the mode-locked pulses is $\sim 150\text{mV}$. Therefore from eqn 4.1(b), the forward current is $\sim 2.65 \times 10^{-10}$ Amps for a $2.5\mu\text{m}^2$ junction. Plugging this into the expression for the forward resistance we get $R_F \approx 1.24 \times 10^8 \Omega$. Also it was estimated in the previous section that the depletion capacitance is $\sim 2.5 \text{ f F}$. Therefore the RC discharge time is $\sim 300\text{nS}$. Also, as the voltage across the capacitance slowly dies away, this fall time will get progressively longer because the forward resistance increases with decreasing forward bias. Conversely even a slightly higher forward bias will exponentially bring down the initial fall time by reducing the R_F .

The measured pulse response shown in Fig 4-22(a) clearly exhibits the long tail that is expected from the above discussion. The fall time is not exactly equal to the RC discharge time because the forward bias in the device might be a little higher than expected (as evidenced by the high photo-response) which in turn means that the forward resistance is lower than the value estimated above.

An interesting experimental consequence of a really long fall time is that if $\tau_F > \tau_{rep}$, where τ_{rep} is the repetition rate of the laser, then the response seen at the output will

start saturating after a few pulses. This is because all the charge separation created by one pulse would not have dissipated by the time the next pulse arrives. The total response would therefore slowly add up until some saturation starts occurring. Due to this the current increase detected by an oscilloscope might be lower than the actual increase that would occur if the incident pulse were indeed a delta function with no repetition. One way to verify if this is happening in the circuit is to directly check for saturation in the photo-response. As was mentioned in section 4.6.3 this was done by lowering the optical power in steps of 3dB and no indication of initial saturation was seen (the response varied linearly with power). A cursory examination of the falling edge of the pulse in Fig 4-22(a) also indicates that this effect is subdued in the detector being measured because the tail almost completely flattens out in $<40\text{ns}$ and the duration between pulses is 50ns . Even in devices where τ_F is indeed greater than τ_{rep} , the degree of saturation should not be that severe. This is because the saturation would be accompanied by a higher induced forward bias across the Ge/Si junction, as the number of neutralized donors in the Si adds up. The higher bias would lower the RC time to a point where all the charge separation created by one pulse would be discharged before the next pulse arrives. Because the RC time varies exponentially with the forward bias, an equilibrium position where there is an equal rise and fall in current during the duration of each pulse, would be reached rather quickly.

Another important problem of having a long fall time that is related to the above discussion but is much more crucial from a practical perspective is obviously Inter-Symbol Interference (ISI) in an actual communication system. Clearly the fall time for

the detector must be less than the bit-rate in the transceiver for the device to be of any use. There are a couple of practical ways of reducing the RC time. The first method is to increase the induced forward bias across the Ge/Si junction for the same amount of incident light, as this will exponentially lower the forward resistance. For a fixed amount of collected photo-charge the induced voltage would be maximum for the lowest amount of gate capacitance (since $V_G \approx Q_{\text{optical}}/C_G$). Without decreasing the device area (which would reduce the amount of light absorbed and also increase resistance), the best way to minimize the capacitance is by reducing the doping on the Silicon channel and thereby increasing the depletion width. Therefore an optimally designed channel can lower the fall time significantly.

A second technique to reduce the tail is to introduce more leakage into the Ge/Si junction, which would directly decrease forward resistance. Therefore, ironically, a worse hetero-interface would aid the performance of the device in terms of the fall time. The leakage could also be increased artificially by adding a tiny metal short across the junction through post-processing. Both of these options, though, would harm the sensitivity of the device because they encourage quicker recombination of the trapped photo-holes in the Germanium.

Even if the problem of a prolonged tail is not completely alleviated by improved device design, there are options available in terms of transceiver circuit modifications that can circumvent the problem to some extent⁵³. Therefore, the slow fall time should not be perceived as an insurmountable roadblock in the use of Photo-HFETs in actual optical communication systems.

4.7 Conclusions: Why use Photo-HFETs?

In this chapter we have proposed and demonstrated a novel Ge/Si photo-detector based on secondary photo-conductivity in a JFET structure. The last few sections have comprehensively dealt with various aspects of the photo-response of the devices including the modeling of the sensitivity and bandwidth using conventional JFET physics. The next question that arises is whether this device design has any crucial advantages over normal detectors that are used in typical communication systems.

The first step towards answering this question was taken in section 4.1.2 where it was shown qualitatively that the increase in channel current in two devices with the same W/L ratios, but completely different areas, would be the same if the same *intensity* of light is incident on both of them. This can also be seen quantitatively from the equation for the Photo-HFET channel current (eqn 4.11). For the same W/L ratio, the rise in I_{Chan} depends only on the change in depletion widths at the source and drain caused by the light, which in turn is proportional to ϕ/A_{Gate} where ϕ is the number of photo-electrons collected and A_{gate} is the area of the Germanium gate (eqn 4.12). Clearly ϕ/A_{Gate} is only dependent on the intensity of the incident light and therefore an equal intensity falling on two gates with equal W/L would cause an identical change in I_{Chan} . Therefore as the device is made smaller *its response stays the same while its capacitance scales down with area. This reduction in capacitance without sacrificing sensitivity is the most important aspect of the Photo-HFET design.*

The overall advantage of this property is worth exploring quantitatively. Consider a device similar to the ones demonstrated in this chapter, but for which the channel thickness, width and length is 100nm (area = $10^{-2} \mu\text{m}^2$). Since the $2.5\mu\text{m}^2$ detector analyzed in section 4.6 showed a change in channel current of a few μA after collecting ~ 2000 photo-electrons, a device with an area of $10^{-2} \mu\text{m}^2$ *would need only ~ 10 photo-electrons to achieve the same change in channel current* (ignoring low-dimensional effects in the smaller detector). This sensitivity even by itself is handy, but coupled with the extremely low capacitance of the device, it becomes extraordinarily useful especially when studied in the context of an actual optical communication system.

Let us suppose that in a wide bandwidth transceiver, 10ps pulses carrying 1000 photons each are incident on both, a standard high speed detector (say a p-i-n photodiode) and a $100\text{nm} \times 100\text{nm} \times 100\text{nm}$ Photo-HFET (average optical power at the input $\approx -19\text{dBm}$, which is a reasonable value for a conventional receiver). Fig 4-26 shows extremely basic circuit diagrams for both configurations. The outputs of the detectors go into very simple open loop trans-impedance preamplifiers with no feedback. These are obviously not ideal circuit designs^{53,54} but will suffice for the purpose of this discussion. The parasitic capacitances, $C_{P,i}$ are primarily composed of the detector (C_D) and wire (C_W) capacitances. In conventional detectors, especially the ones flip-chip bonded to typical transceiver circuits, the length of the wire and that of the active region of the detector easily adds up to $>100\mu\text{m}$. If the other two dimensions (thickness and width) are taken to be equal, then the 100 microns of semiconductor will have a capacitance of $\sim 10\text{fF}$ (this is

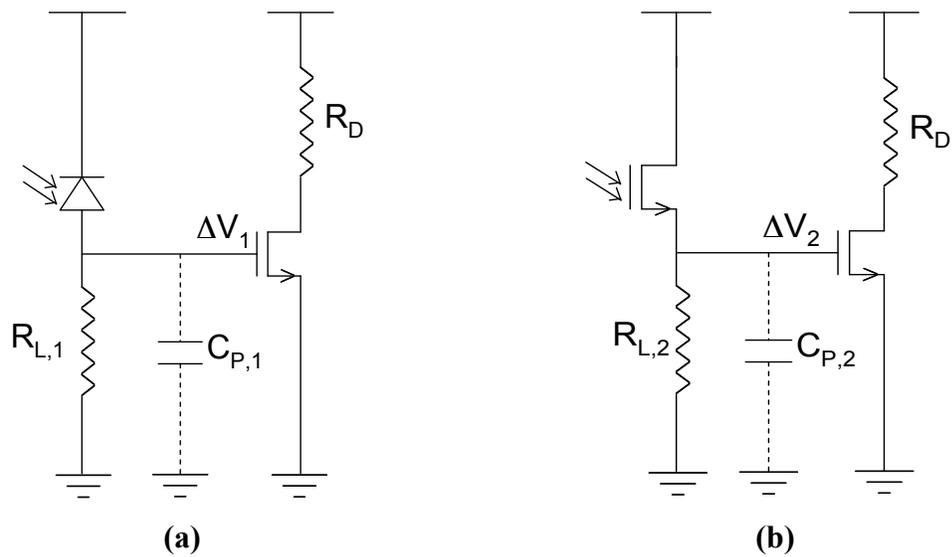


Fig 4-26: Simple open loop trans-impedance amplifier circuit for: (a) Standard p-n junction photo-detector (b) Photo-Heterojunction FET. The parasitic capacitance of the Photo-HFET is much smaller than that of the other device. Therefore $\Delta V_2 \gg \Delta V_1$, if an equal amount of charge is generated by the light in both cases.

in fact a rather optimistic estimate for the capacitance seen in most practical detectors^{55,56,57}). A 100nm x 100nm x 100nm Photo-HFET on the other hand will have a capacitance of only 10 aF. Also *since there is no wire on the Ge gate*, the wire capacitance comes only from the connection to the preamp transistor, which in a highly integrated circuit will be on a length scale similar to the transistor itself. Therefore the length of that wire can also be expected to be of the order of 100nm. Hence *the total parasitic capacitance in the HFET preamp circuit is only a few tens of atto-Farads, i.e., nearly 1000 times less than that in a conventional p-n junction detector*. It is worth

repeating that this huge reduction in capacitance greatly enhances the utility of the proposed devices and this fact is clearly illustrated by the remainder of the analysis below.

Returning to the transceiver circuits, the input resistance for each pre-amp (i.e., the load resistance for each detector) is chosen so that $R_{L,i} C_{P,i} = 1/\omega_B$ where ω_B is the required bandwidth in the receiver. Therefore in both circuits, the voltage induced at the gate of the gain transistor is:

$$\Delta V_i = I_{Photo} R_{L,i} = \frac{I_{Photo}}{\omega_B C_{P,i}}$$

where, I_{Photo} is the photo-current at the output of the detectors. Since $\omega_B \approx 1/\tau_{Pulse}$, where τ_{Pulse} is the pulse width, we get:

$$\Delta V_i \approx \frac{I_{Photo} \tau_{Pulse}}{C_{P,i}} = \frac{Q_{Photo}}{C_{P,i}} \quad (4.19)$$

Using eqn 4.19 a comparison can now be made between a conventional high speed detector and a Photo-HFET. In a regular photo-diode, all 1000 of the incident photo-electrons can ideally be collected from each 10ps pulse (assuming 100% quantum efficiency). Therefore the total charge, $Q_{Optical}$, flowing into the trans-impedance amplifier circuit is $1000 \times 1.6e-19 = 1.6 \times 10^{-16}$ C.

For the Photo-HFET, as was mentioned before, only ~10 collected photo-electrons can change the channel current by a few μ A. Therefore even if the Ge gate absorbs only 1% of the incident 1000 photons, the output current would be sufficiently high. Of course in the measurements on our devices, the ratio of the absorbed to the

incident power is only $\sim 10^{-5}$. But that was because nothing was done to better confine the incoming light or to improve the absorption cross-section of the Ge gate. An optimized waveguide based resonator structure, like the one depicted in Fig 4-5, should readily achieve an absorption efficiency of 1% if not more. In any case, assuming that only 10 out of the 1000 incident photons are absorbed gives a significant amount of output photo-current. Also, since the device is extremely small, its intrinsic rise time will be of the order of 1ps (see section 4.6.4) and therefore the output current will indeed rise to its maximum value through the duration of a 10ps pulse. (This rather straightforward observation is actually very crucial. If the rise time of the detector is of the order of the pulse width, then the amount of charge that flows in the external circuit in the presence of the light will only be equal to the number of photo-electrons collected by the gate, which is of course extremely small due to the small size of the gate. In other words there wouldn't be any useful current gain which in turn would significantly reduce the effectiveness of the device).

Going back to the calculation of Q_{Optical} for the Photo-HFET, let us assume for ease of comparison that $\sim 1.6\mu\text{A}$ of average output current flows for the duration of the pulse. And so we get $Q_{\text{Optical}} = 1.6\mu\text{A} \times 10\text{ps} = 1.6 \times 10^{-17} \text{ C}$ (i.e., the equivalent of 100 electrons). Therefore we see that the total charge at the input of the gain transistor is 10 times greater for the conventional p-n junction detector.

Substituting the values of the capacitance and the total charge for both devices into eqn 4.19, we see that *the voltage at the pre-amp created by the Photo-HFET is ~ 100 times that of a typical junction detector*. Therefore the 1000 fold reduction in capacitance

completely dominates over the ten-fold reduction in photo-charge and hence, in stark contrast to the $\sim 10\text{mV}$ produced by the standard device, the HFET induces $\sim 1\text{V}$ at the gate of the gain transistor thereby providing enough voltage to completely switch it.

Taking this a step further, if we assume that the resonator around the Ge gate ensures that almost all the incident light is absorbed, then theoretically, the Photo-HFET will give the same output voltage *with only 10 photons per incoming pulse*. Such response would represent *a four orders of magnitude enhancement over the performance of a conventional detector*.

The increase in the induced voltage described above *directly translates into a vast improvement in detector sensitivity*, because very little optical power at the input would still achieve the signal-to-noise ratio needed to conform to the bit error rate specifications of the system⁵⁸.

4.7.1 Conclusion

The above discussion shows quantitatively that the tiny capacitance associated with the proposed photo-detectors plays a crucial role in determining the overall performance of the device in a transceiver circuit. As all the values used in the calculations are typical for standard receivers, it can be safely inferred that the Photo-HFET structure can provide huge improvements in sensitivity over discrete and even other integrated detectors. The only unknown is the design of the cavity structure needed to increase the absorption of light by the Ge gate, but even that shouldn't pose a very significant challenge since an external quantum efficiency of only 1% is sufficient to provide impressive gains in overall performance. In fact even without the resonator,

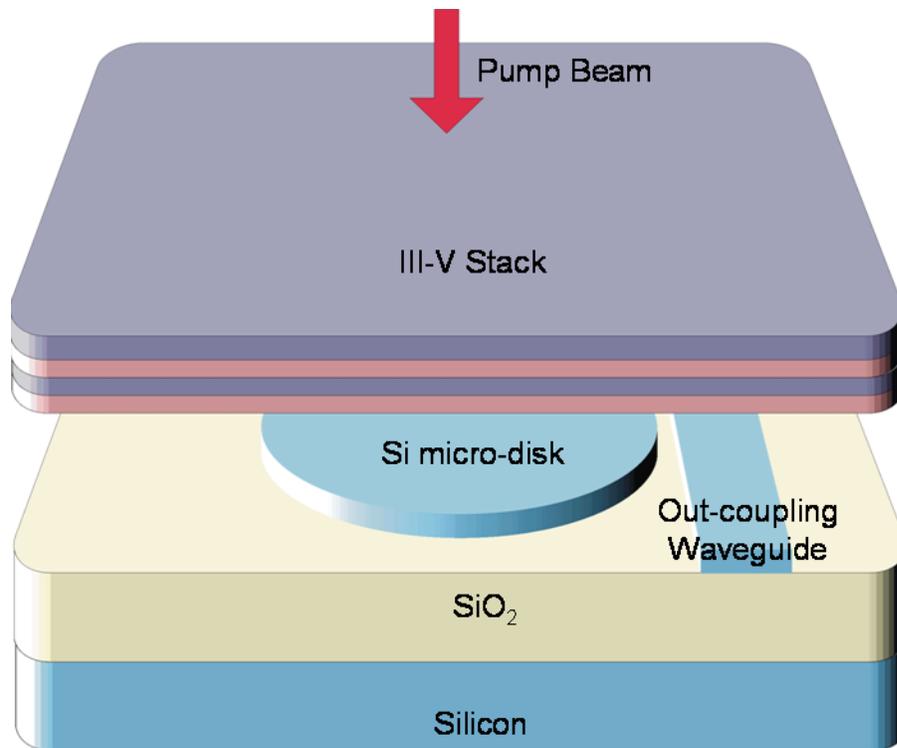
slightly better quality Germanium should be enough to produce at least a ten-fold enhancement in sensitivity over conventional detectors. Therefore, in conclusion, the Photo-HFET represents a highly scalable Germanium detector for Silicon photonic applications that has the potential to significantly improve the performance parameters of very high bandwidth Si receiver circuits.

Chapter 5: Design of a Hybrid III-V/Silicon Evanescent Laser

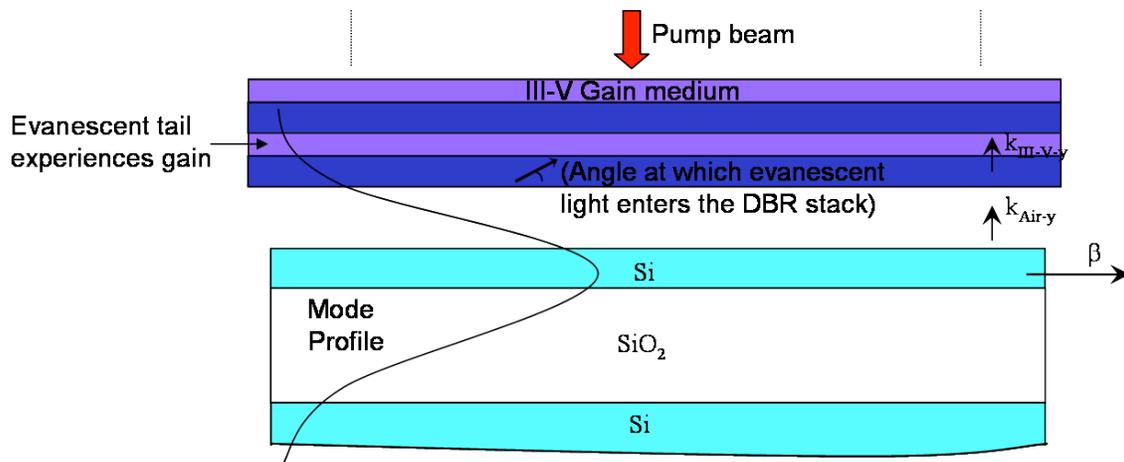
In the introductory chapter, the advantages of having a $1.55\mu\text{m}$ laser integrated onto a Silicon chip were discussed in the context of a complete Si photonics platform. As was mentioned there, several devices have been reported in the literature that take different approaches towards solving this problem, with varying degrees of success. Here, we take a look at a specific laser geometry that is based on a hybrid structure on Silicon that utilizes III-V alloys. First, the fundamentals of the device are outlined based on basic physical principles. Then, various optical properties of the design are explored in detail using *finite-element* electromagnetic modeling. These simulations provide a feasibility analysis for this laser structure and give a framework based on which practical devices can be fabricated. Finally we summarize preliminary experimental efforts towards the realization of a hybrid III-V/Si laser.

5.1 *Basic design*

As has been discussed in earlier chapters, the indirect band-gap of Silicon precludes its use as the gain medium in a laser. A III-V/Si hybrid configuration solves this problem by integrating a III-V material into a Silicon cavity in order to provide gain



(a)



(b)

Fig 5-1: (a) Schematic of a III-V stack evanescently coupled to a Si micro-disk. The waveguide can be used to couple the light out of the laser. (b) Cross-sectional diagram of the device. The light inside the Si will enter the III-V at an angle determined by the effective index of the disk mode

at the appropriate wavelength (which in a communications context is $1.55\mu\text{m}$). The integration of the III-V alloy can be done in various ways depending on the overall gain requirements of the system. In the proposed device that will be analyzed in this chapter, a III-V stack is *evanescently coupled* to a Silicon microdisk (Fig 5-1). The motivation for using this geometry comes from recent demonstrations of extraordinarily high quality factor (Q) cavities on SOI. Microdisk resonators with Q values far exceeding 10^6 have been reported in the literature^{21,22}. The main consequence of this improvement is that the loss in the cavities is reduced to remarkably low levels (for a Q of 10^6 the loss is of the order of 0.5dB/cm^{59}). Therefore in the context of making a laser, a very small amount of gain is needed to overcome the cavity losses and reach threshold.

Keeping this in mind the structure of the device shown in Fig 5-1 can be better understood. Assuming that the microdisk has a high enough Q, *the gain experienced by just the evanescent tail in the III-V is enough to achieve lasing*. Hence, a laser can be constructed where the light stays almost completely confined in the Silicon and therefore benefits from the exceptional loss characteristics of the microdisk. Consequently the lasing mode is essentially a mode of the Si disk with the III-V appearing only as a perturbation.

5.1.1 Design of the III-V stack: Reflection along with gain

Now that the fundamental idea behind the proposed device has been introduced, the details of the design of the III-V gain medium can be established. Since the III-V alloys are also high refractive index materials, *the light in the Silicon disk will tend to escape into a bulk III-V substrate if it is brought close enough to access the evanescent*

part of the mode. This can be seen as follows: if the effective index of a mode in a Silicon slab with air cladding is n_{eff} , then the transverse wave-vector in the air will be (Fig 5-1(b)):

$$k_{Air-y} = \frac{\omega}{c} \sqrt{1 - n_{eff}^2} \quad (5.1)$$

Since this is imaginary, it represents the exponentially decaying tail of the mode. Now, if the tail enters a high index III-V alloy that has a refractive index that is greater than the modal index then the transverse wave-vector in the III-V is given by:

$$k_{III-V-y} = \frac{\omega}{c} \sqrt{n_{III-V}^2 - n_{eff}^2} \quad (5.2)$$

where, it is assumed that the presence of the III-V does not significantly alter the profile of the mode (the effective modal index is still n_{eff}). Now, if $n_{III-V} > n_{eff}$, then $k_{III-V-y}$ is real, i.e, the optical field does not decay transversely in the III-V and can effectively radiate away.

This loss of light into the III-V obviously completely ruins the Q of the cavity and has to be eliminated. One way of doing this is to ensure that the III-V layers are very thin so that they do not support a continuum of modes into which the light can couple. This can be accomplished by wafer-bonding an epitaxial stack of the alloys onto the Silicon and then removing the substrate. Unfortunately in such a bonded configuration the III-V and the Si are in close proximity (1-2 nm) and therefore the coupling between the two is no longer evanescent. The overall laser mode is a 'super-mode' of the III-V and the Si with a considerable fraction of the power being confined in the former. The disadvantage

of this design is that the excessive overlap with a number of epitaxially grown layers can add losses due to which the exceptional Q of the disk is compromised. Also the atomically close wafer bonding process might not be very easily integrable with a conventional CMOS process. Therefore, even though the demonstrated devices using wafer-bonding^{60,61,62} show considerable promise, it is prudent to look at other techniques to reduce the losses through the III-V substrate.

In our work, we use an *innovative layered structure* in the gain medium to keep the light confined predominantly in the Silicon. In addition to providing gain at 1.55 μm , the III-V is designed as a *Distributed Bragg Reflector (DBR) in the same wavelength window*. An important feature of the DBR is that it is not designed for normal incidence: *the stack layers are chosen so that the optimum incident angle for reflection is the angle that the evanescent portion of the mode makes in the III-V* (Fig 5-1(b)). If a sufficient amount of reflection is obtained, then the light will stay in the Silicon in spite of the close proximity of a high index gain medium.

The advantages of using this hybrid DBR/microdisk approach mainly lie in the ease with which such devices can be integrated onto a Silicon chip. As will be seen later in this chapter, if the Q of the microdisk is high enough, the spacing between the III-V and the Silicon doesn't have to be controlled that accurately in order to achieve lasing. This in turn greatly relaxes the constraints on the wafer bonding procedure^{63,64} that adheres the DBR stack to the Silicon, thereby making it a lot more compatible with a normal CMOS foundry process. This advantage is further enhanced by the fact that after the bonding *the III-V substrate doesn't have to be removed, due to the presence of the*

DBR. The reduction in post processing created by skipping the substrate etch step is extremely significant in terms of the integrability of a hybrid III-V/Si laser into CMOS. Finally, the use of a normal semiconductor gain medium means that *the laser can be electrically pumped*, which is an important requirement for an integrated light source that has to be used in an actual high bandwidth transceiver.

It is worth clarifying that the use of the term ‘evanescent’ inside the DBR is not strictly correct since the mode does not exponentially die off in the III-V’s owing to their high refractive index. The field is actually evanescent only in the space between the DBR and the Silicon. Nonetheless, since this distinction is quite obvious in the context of this device, the ‘evanescent’ terminology is used numerous times in this chapter for ease of description.

5.1.2 Calculation of DBR layer thicknesses

As has already been mentioned in the previous section, the most crucial aspect of designing the DBR is that its optimum reflection angle should be the *evanescent* angle in the III-V of the overall SOI mode. In other words instead of the overall propagation constant of the mode, *the DBR is designed for the transverse wave-vectors in the III-V* (shown in Fig 5-1(b) and expressed in eqn 5.2). From the wave-vectors, the appropriate thicknesses can be calculated by using the standard quarter wave approximation⁶⁵. Therefore the thickness of the i^{th} layer is given by:

$$t_i = \frac{1}{4} \left(\frac{2\pi}{k_{\text{III-V-}y}, i} \right) \quad (5.3)$$

This stack configuration would give perfect phase cancellation in the transmitted wave for the evanescent portion in each III-V layer.

From eqns 5.2 and 5.3 it can be seen that the effective index of the mode in the Si is needed to extract the values of the DBR layer thicknesses. Coming up with an exact value for the n_{eff} of a disk resonator mode involves solving the cylindrical form of Maxwell's equations⁶⁶, which is rather complicated. Fortunately the high index contrast between Si and air/oxide strongly confines light in the micro-disk in the vertical direction thereby reducing the problem to 2 dimensions⁶⁶. Under this approximation the transverse k-vector of the disk mode, $k_{\text{Si-y}}$, is completely determined by the Si thickness. The *total* in-plane propagation constant (β) is therefore simply the *slab index* of the Si layer and is always equal to:

$$\beta = \sqrt{\left(\frac{2\pi n_{\text{Si}}}{\lambda_0}\right)^2 - k_{\text{Si-y}}^2} \quad (5.4)$$

β can now be easily converted into an n_{eff} ($n_{\text{eff}} = \beta\lambda/2\pi$), which can be plugged into the equations for the DBR thicknesses (5.2 and 5.3). The beauty of this method is that, if the lasing wavelength is known, n_{eff} can be estimated from just the thickness of the top Silicon layer without specifying the radius of the disk or the radial order of the mode in question. *The direction of β will vary according to the radial and angular position on the disk*⁶⁷, but its magnitude will be roughly equal to that given by eqn 5.4.

In the context of finding the ideal DBR layer thicknesses it must be noted that the effective index used in the calculations will itself *be altered by the interaction of the DBR with the disk mode*. But to get a first order estimate, the approximation that the effective

index of the Si slab mode remains unperturbed by the DBR, is reasonable. To get a more accurate picture, the reflectivity of the stack designed in this manner is explored in greater detail in the numerical analysis described later in this chapter. The finite element simulations take into account the effect of the DBR-Si modal interaction on the performance of the DBR.

5.1.3 Material system for the DBR

There are a number of important considerations that have to be taken into account before picking the material system for building the DBR. Firstly, the epitaxial growth of the DBR should be viable on a standard substrate (like InP or GaAs) in order to provide a feasible platform on which to fabricate the stack. Secondly there should be at least one material in the stack that can provide gain at the required wavelength (which in this case is 1.55 μm). Thirdly the band-gap distribution in the DBR has to be such that the excess carriers created by the pump stay confined inside the gain medium. Finally, it is imperative that there be sufficient index contrast between the films used in the DBR so that a very thick stack can be avoided and significant amount of reflection can be achieved with a reasonably small number of layers.

Keeping these requirements in mind, *the novel AlGaAsSb system* was chosen for designing and fabricating the stack. These are quaternary alloys of AlAs, AlSb, GaAs and GaSb that can have a vast variation in band-gap (and therefore in refractive index) depending on the concentration of Al and Ga⁶⁸. As seen in Fig 5-2, Al and Ga rich alloys can be grown lattice matched to InP: the appropriate concentrations are AlAs_{0.56}Sb_{0.44} and

GaAs_{0.51}Sb_{0.49}. The band-gaps for these alloys are $\sim 2.0\text{eV}$ and $\sim 0.8\text{eV}$ respectively (AlAsSb is indirect and GaAsSb is direct) and the refractive indices are ~ 3.1 and 3.75 (the exact lattice constant, band-gap and refractive index for each composition depends on the bowing parameters⁶⁸ for the various alloys. Appendix A lists the methodology used in the calculations). The 0.8eV gap exactly corresponds to the $1.55\mu\text{m}$ window and therefore the Ga rich alloy is ideally suited to be the gain medium. Additionally the alignment between AlAsSb and GaAsSb is type I^{68,69} and a significant part of the 1.2eV difference in band-gap is seen between the valence and conduction bands thereby ensuring

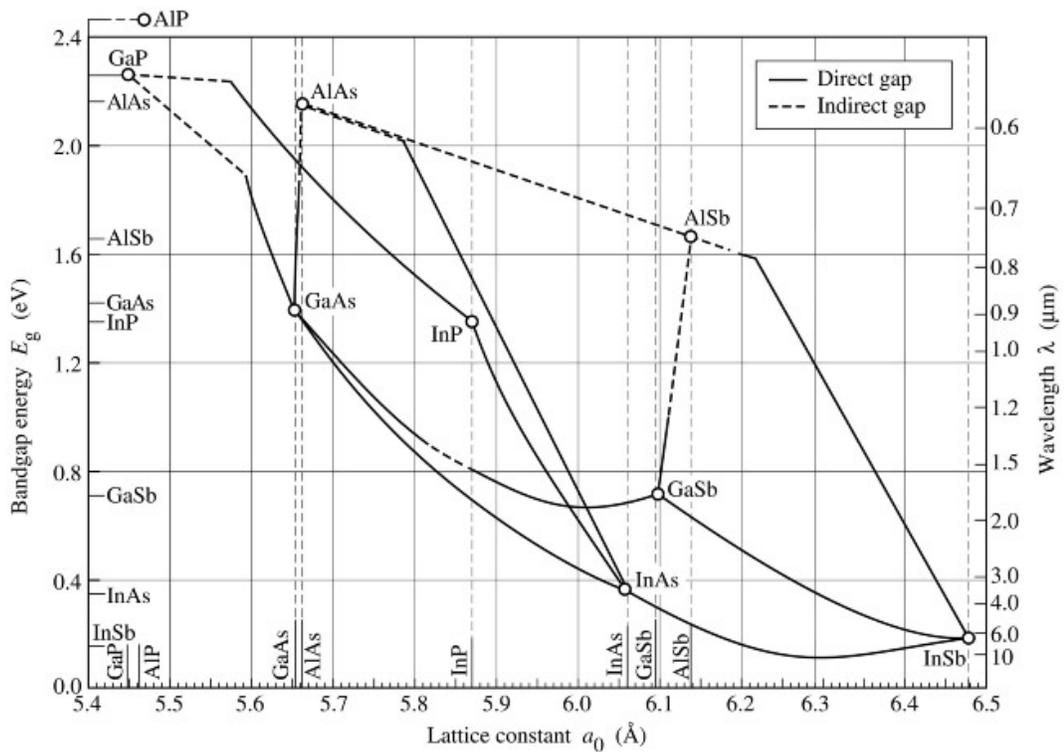


Fig. 7.6. Bandgap energy and lattice constant of various III-V semiconductors at room temperature (adopted from Tien, 1988).

Fig 5-2: Dependence of bandgap on lattice constants for standard semiconductors

that carriers pumped into the gain region will remain confined there. And most crucially, the index contrast (~ 0.65) is much higher than what can be achieved from a more conventional material system in the same wavelength window (InGaAsP based alloys can give a contrast of ~ 0.2).

It must be mentioned here that an important factor affecting the final design of the DBR is the index dispersion in the gain layers. It is well known that the refractive index of any semiconductor is highly dispersive at wavelengths close to its band-edge⁷⁰ and therefore the index of the GaAs_{0.51}Sb_{0.49} alloy will be extremely sensitive to the exact lasing wavelength (λ_L) in the 1.55 μ m window. Unfortunately the exact value of λ_L will be determined by the resonance conditions inside the micro-disk and by the position of the resonant wavelengths within the gain spectrum of the Ga rich alloy. This uncertainty in λ_L means that the index of the GaAs_{0.51}Sb_{0.49} layers cannot be estimated accurately, which in turn precludes their use as reflecting layers throughout the DBR. Therefore in our design, in the majority of the DBR they are replaced with Al_{0.15}Ga_{0.85}As_{0.52}Sb_{0.48} layers that have a band-gap of ~ 1.1 eV and a refractive index of ~ 3.55 . These values still provide sufficient index contrast while minimizing dispersion in the 1.55 μ m window.

Plugging the above parameters into eqn (5.3), the final thicknesses for the DBR are calculated. n_{eff} is assumed to be 2.5, which is the slab index of the TE₀ mode for a 150nm thick Si film. The Si thickness comes from the SOI wafers that were used in proof-of-concept experiments described in section 5.3.2. Fig 5-3 shows a schematic of the final structure: the GaAs_{0.51}Sb_{0.49} gain layers are at the top of the stack and will obviously be closest to the disk resonator. Only 2 layers are chosen with that composition

because finite element modal calculations estimate the fill factor to be sufficient to overcome the expected losses in the micro-disk (the simulations are described later in the chapter). By limiting the number of gain layers to the minimum required, the index dispersion is restricted to as low a value as possible. In addition, even though the fill factor would be maximized if the top most film (closest to the disk) is a gain layer, excess surface recombination of the pumped carriers would render that configuration useless. Hence the top-most layer is the large band-gap, low index $\text{Al}_{0.95}\text{Ga}_{0.05}\text{As}_{0.56}\text{Sb}_{0.44}$ alloy (a small amount of Ga is added to the alloy to make it more stable in air).

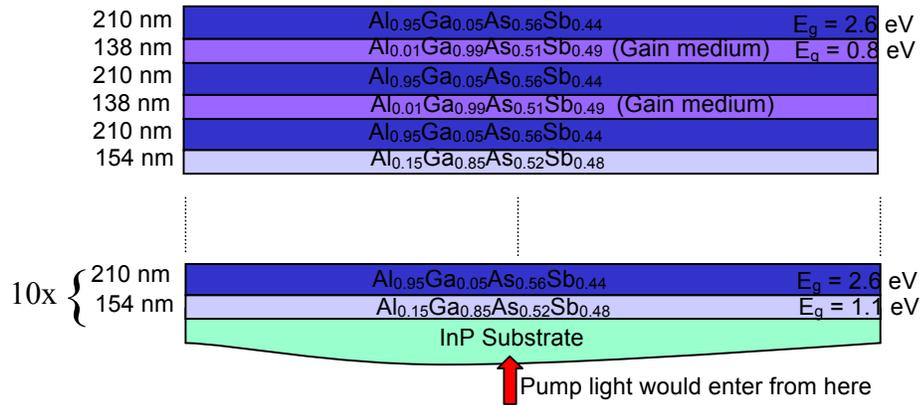


Fig 5-3: Final design for the DBR. The gain layers are limited to the top section of the stack that will be closest to the micro-disk.

It is also worth mentioning that the large band-gap of the $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As}_{0.52}\text{Sb}_{0.48}$ layers, allow the 2 gain layers at the top to be pumped optically from the back, with any wavelength $> 1.1\mu\text{m}$ (and $< 1.55\mu\text{m}$). That is relevant since any wavelength in that window will not be absorbed by Si either and will therefore not cause any unnecessary free carrier generation in the Si disk which in turn can reduce its Q. Therefore in

conclusion, the structure depicted in Fig 5-3 can produce gain @1.55 μm (with electrical or optical pumping), while providing reflection in the same wavelength window. When used in conjunction with a high Q Silicon disk through evanescent coupling, this DBR represents a unique solution for producing integrated lasers on Si. The next section looks at simulations that predict the optical and electrical behavior of the structure when it is brought into close proximity of a micro-disk.

5.2 Numerical Analysis

5.2.1 Transfer matrix formulation

The goal of the numerical analysis is to ascertain the losses for an optical mode in a micro-disk that is brought in close proximity of the DBR stack. The first step is to estimate the reflectivity of the DBR for light that is incident on it at the evanescent angle determined by the effective index of the mode in the Silicon. This can be done by invoking basic wave-transfer matrix theory⁶⁵ that is commonly used to find the reflectivity of DBRs. The transfer matrices inside layer m (P_m) and at interface n (T_n) in the DBR are given as:

$$P_m = \begin{bmatrix} e^{i k_{III-V-y} t_{III-V}} & 0 \\ 0 & e^{-i k_{III-V-y} t_{III-V}} \end{bmatrix} \quad (5.5)$$

$$T_n = \frac{1}{t_{21}} \begin{bmatrix} t_{21}t_{12} - r_{11}r_{22} & r_{11} \\ -r_{22} & 1 \end{bmatrix}$$

where, referring to Fig 5-4,

$$\begin{bmatrix} E_{a2} \\ E_{b2} \end{bmatrix} = P_m \begin{bmatrix} E_{a1} \\ E_{b1} \end{bmatrix} \quad \begin{bmatrix} E_{11} \\ E_{21} \end{bmatrix} = T_n \begin{bmatrix} E_{12} \\ E_{22} \end{bmatrix}$$

In eqn 5.5, $t_{\text{III-V}}$ is the thickness of the III-V layer in question and the t_{ij} and r_{ij} terms are the reflection and transmission coefficients at each interface that are expressed in terms of refractive indices as:

$$\begin{aligned} t_{12} &= \frac{2n_1}{n_1 + n_2}, & t_{21} &= \frac{2n_2}{n_1 + n_2} \\ r_{11} &= \frac{n_1 - n_2}{n_1 + n_2}, & r_{22} &= \frac{n_2 - n_1}{n_1 + n_2} \end{aligned} \quad (5.6)$$

From the above equations we find that $t_{21}t_{12} - r_{11}r_{22}$ is equal to 1.

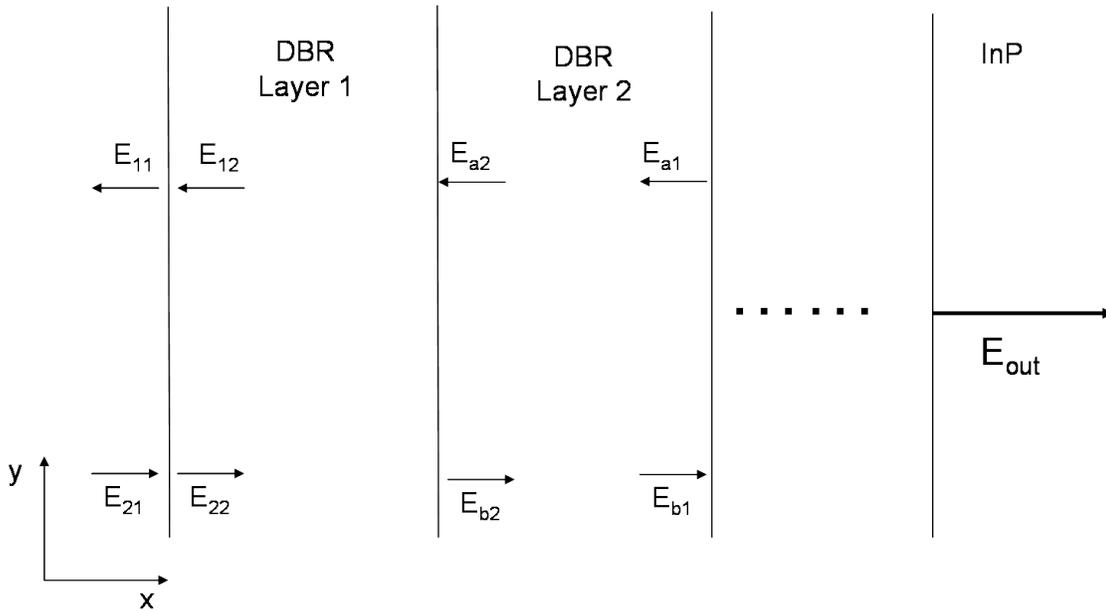


Fig 5-4: The DBR structure (with the layers going from left to right). The fields on both sides of an interface and at the two ends of a layer are depicted. The propagation and transmission matrices in eqn 5.5 give the relationships between the fields. The light that is not reflected by the DBR and escapes through the InP is represented by E_{out} .

The matrices in eqn 5.5 can be multiplied for all the layers and interfaces in the stack to find the transfer matrix between the electric field at the two ends of the DBR. Now, from Fig 5-4 we can see that if the DBR is grown on an InP substrate, there is no x component at the end of the stack and therefore the relation between the field at the input and output of the stack is as follows:

$$\begin{bmatrix} E_{11} \\ E_{21} \end{bmatrix} = M \begin{bmatrix} 0 \\ E_{Out} \end{bmatrix} \quad \text{i.e.,} \quad \begin{bmatrix} E_{11} \\ E_{21} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{21} \\ M_{12} & M_{22} \end{bmatrix} \begin{bmatrix} 0 \\ E_{Out} \end{bmatrix} \quad (5.7)$$

where, M is the product of all the individual transfer matrices. Since E_{21} is the field incident on the DBR and E_{12} represents the overall reflected field, *their ratio gives us the reflection coefficient of the stack*. Hence, expanding eqn 5.7, we see that the reflectivity of the DBR is $\sim |M_{21}/M_{22}|^2$.

An important feature of the calculation is that since the angle of incidence is not 90° , the reflectivity at each interface, r_{ij} , is no longer polarization independent and for the TE mode is higher than what would be expected for normal incidence. The higher value of r allows us to use fewer layers in the DBR to achieve the same amount of reflection. The way this manifests itself in the above formulae is that the n_i 's used to calculate r at each interface (eqn 5.6) are not the indices of the III-V films but the *effective indices of the transverse component of the wave-vector in each III-V layer* ($n_{III-V,y} = \lambda_0 k_{III-V,y}/2\pi$). This makes sense because the transverse component of the incident light is the one that gets reflected by the DBR, which is also why the wave-vector in the thickness estimate in eqn 5.3 and in the exponential phase terms in 5.5 is also $k_{III-V,y}$. In the current device, the contrast between the transverse indices is higher than that between the actual indices of

the two III-V films (~ 0.7 vs ~ 0.45), which gives the higher reflection predicted by the oblique incidence.

The value of $k_{\text{III-V-y}}$ for each layer is given by eqn 5.2, which in turn needs the in-plane effective index of the Si mode, n_{eff} . As described earlier, the in-plane propagation constant inside the disk depends on the thickness of the Si layer and the exact wavelength of the mode inside it (the lasing mode in this context). n_{eff} was taken to be 2.5 which was the same value used in the calculations for the DBR. The wavelength window is taken to be around 1550nm as usual. Plugging these values into eqns 5.5 to 5.7 along with the layer thicknesses and indices used in the actual DBR, we can estimate its reflectivity.

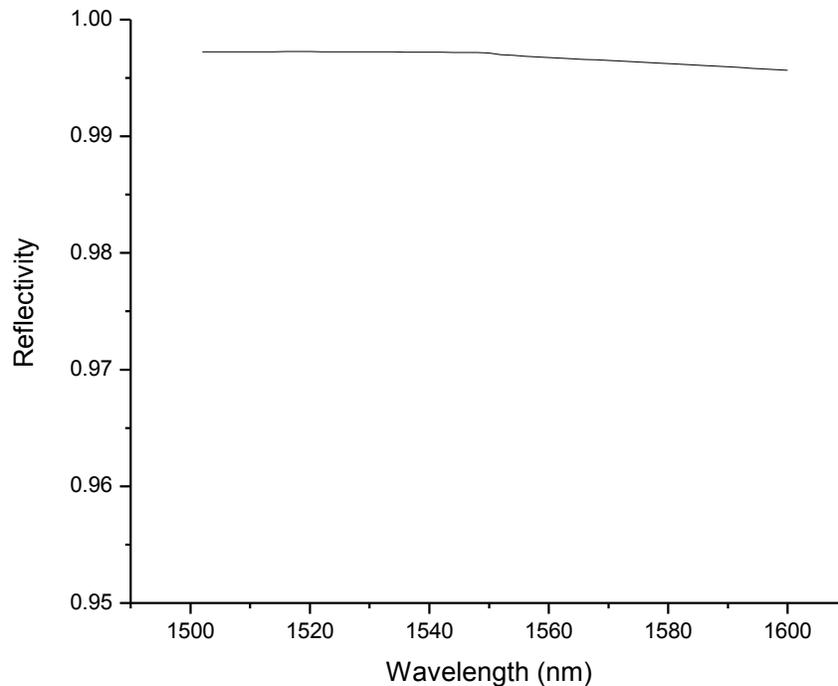


Fig 5-5: Estimated reflectivity for the DBR at different wavelengths. The top Si thickness is assumed to be 150nm (giving an effective Si slab index of 2.5 at 1.55 μm)

Fig 5-5 shows the estimated reflectivity for a wide wavelength range. Material and modal dispersion was taken into account in the Silicon in order to accurately calculate n_{eff} at each λ . Dispersion in the III-V layers was also included which gives the slight kink at $1.55\mu\text{m}$. It can be seen that the reflectivity of the DBR is $>99\%$ over a fairly wide wavelength window. This is crucial even though the DBR is designed to work at $1.55\mu\text{m}$, because as mentioned in an earlier section there can be *shifts in the lasing wavelength due to perturbations caused to the disk mode by the DBR itself*.

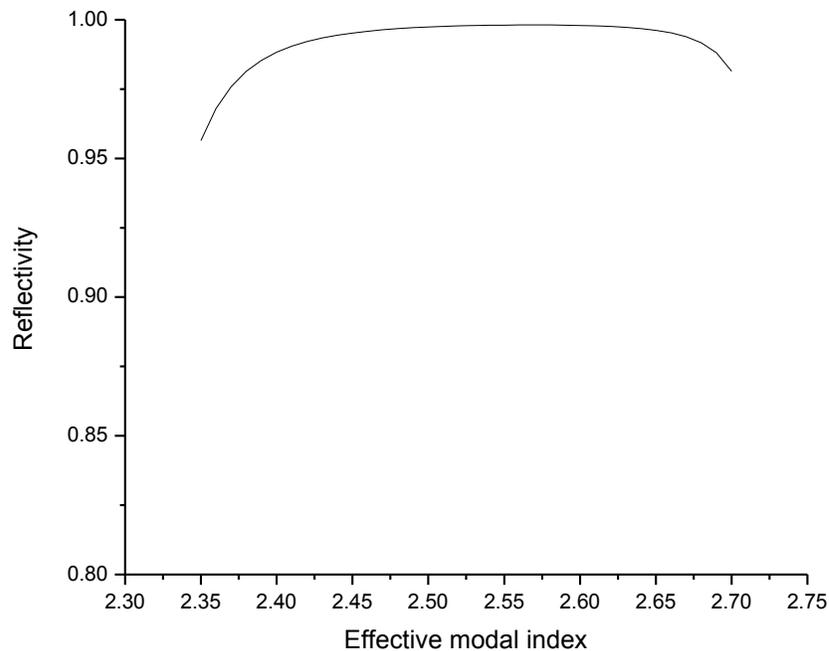


Fig 5-6: Estimated reflectivity for the DBR at different effective indices for the Si slab mode. The wavelength is fixed at $1.55\mu\text{m}$

Fig 5-6 shows the reflectivity at $1.55\mu\text{m}$ for varying n_{eff} . The aim of this plot is to see whether the DBR can accommodate changes in n_{eff} that can again be due to the proximity

of the DBR to the Si mode. It can be clearly seen that the DBR shows a wide window of operation, indicating that the design is robust in both the wavelength and effective index parameter spaces. In fact even though the DBR is designed for $n_{\text{eff}} \sim 2.5$, the actual peak in the curve occurs at a higher index because that represents a larger incidence angle, which in turn results in greater reflection from all the interfaces.

In the next section, we take a more detailed look through finite element simulations, at the modal characteristics of a Si micro-disk evanescently coupled to the DBR.

5.2.2 Finite Element analysis

The next step towards understanding the optical behavior of the device is to find the overall modes of the disk/DBR structure and to investigate the loss mechanisms for those modes. A commercial software package, Lumerical Mode Solutions, is used for this purpose. It is a full vectorial finite element mode solver made primarily for straight waveguides with arbitrary cross-sections. In addition it has the ability to add curvature to the waveguides so that micro-disk behavior can be emulated.

The device is first simulated in 1-d in order to see the efficacy of the DBR and to find the electric field profile of the fundamental TE ‘super-mode’ of the structure. The exact thicknesses of the stack and the top Si layer in the SOI are used; the only available variable is the separation between the two and that is taken to be 100nm. As will be seen later in this section, the trade-off between disk gain and loss is not overly sensitive to this parameter and so 100nm was settled on to make sure that there is sufficient overlap of the Si mode with the DBR, thereby making the simulations relevant. This separation is main-

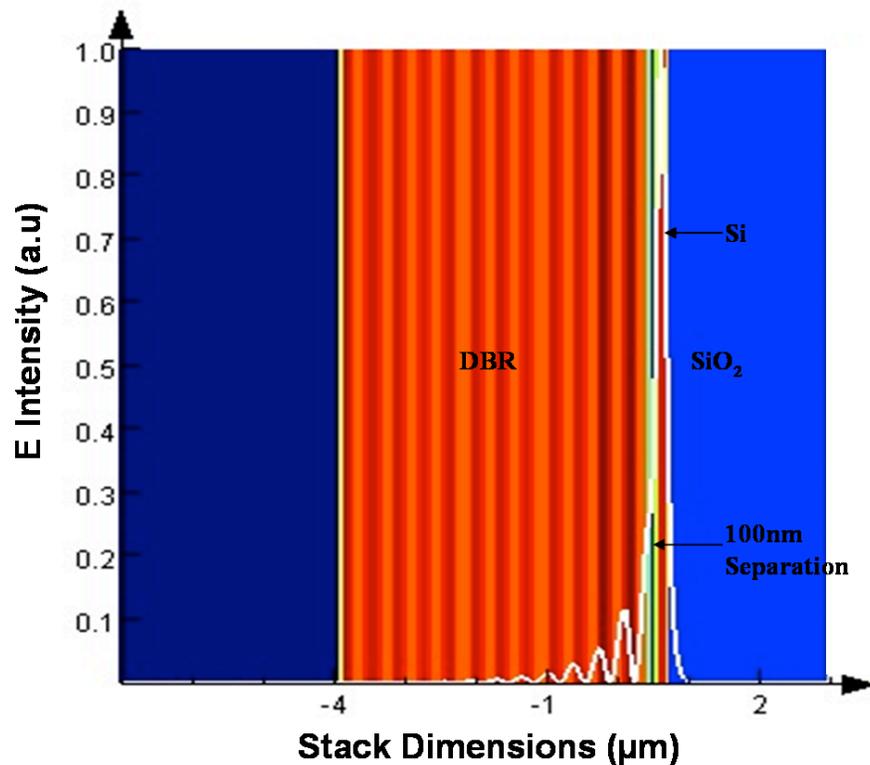


Fig 5-7: Results of a 1-d Finite Element simulation of the DBR coupled to a Si slab. The separation between the two is 100nm. The decaying electric field profile indicates the reflection in the stack

tained in all the subsequent 2-d simulations as well and serves as a baseline design parameter for making a hybrid Si/III-V laser.

The results of the 1-d analysis at $1.55\mu\text{m}$ are shown in Fig 5-7. The decaying profile inside the DBR perfectly illustrates its effectiveness as a reflector for the evanescent portion of the mode in the Silicon. The effective index of the overall mode is 2.58 compared to ~ 2.5 for the original Si slab mode showing the perturbation caused by the DBR.

Now that the presence of a ‘super-mode’ in the structure has been established, the 1-d calculation can be extended to two and three dimensions in order to get a better picture of the losses involved and the fill factor within the gain layers. The main loss mechanisms in the device are light transmission through the III-V substrate (i.e., the light that the DBR doesn’t reflect) and bending losses. The losses through the substrate are crucial even though the reflectivity of the DBR is estimated to be >99%. This is because if the Si waveguide is brought next to the III-V in the absence of the DBR the losses will be of the order of 10^4 /cm. The bending losses occur because *there is no lateral confinement for the part of the mode inside the DBR*. As was explained earlier, the light inside the DBR is not actually evanescent, due to the high indices of the III-V layers. Therefore as the disk is made smaller the relative strength of coupling to the slab modes inside the III-V becomes higher thereby significantly increasing bending losses.

Regarding the 2-d and 3-d calculations it can be seen that since the software finds the modes in a waveguide, the modes of a device involving a disk cannot be extracted directly. Those are therefore solved for as follows: first a rectangular waveguide is created in 150nm thick Silicon that has an n_{eff} of ~ 2.5 (see the top section of Fig 5-8). The width of the waveguide is picked so that it is large enough for the sidewalls to not interact much with the mode but is small enough to be computationally viable. In our simulations, the width was taken to be $5\mu\text{m}$. This waveguide is then brought within 100 nms of the DBR to give us the starting structure (Fig 5-8). *The next step is to add curvature to the waveguide with a radius equal to that anticipated for a practical micro-disk*. This is done

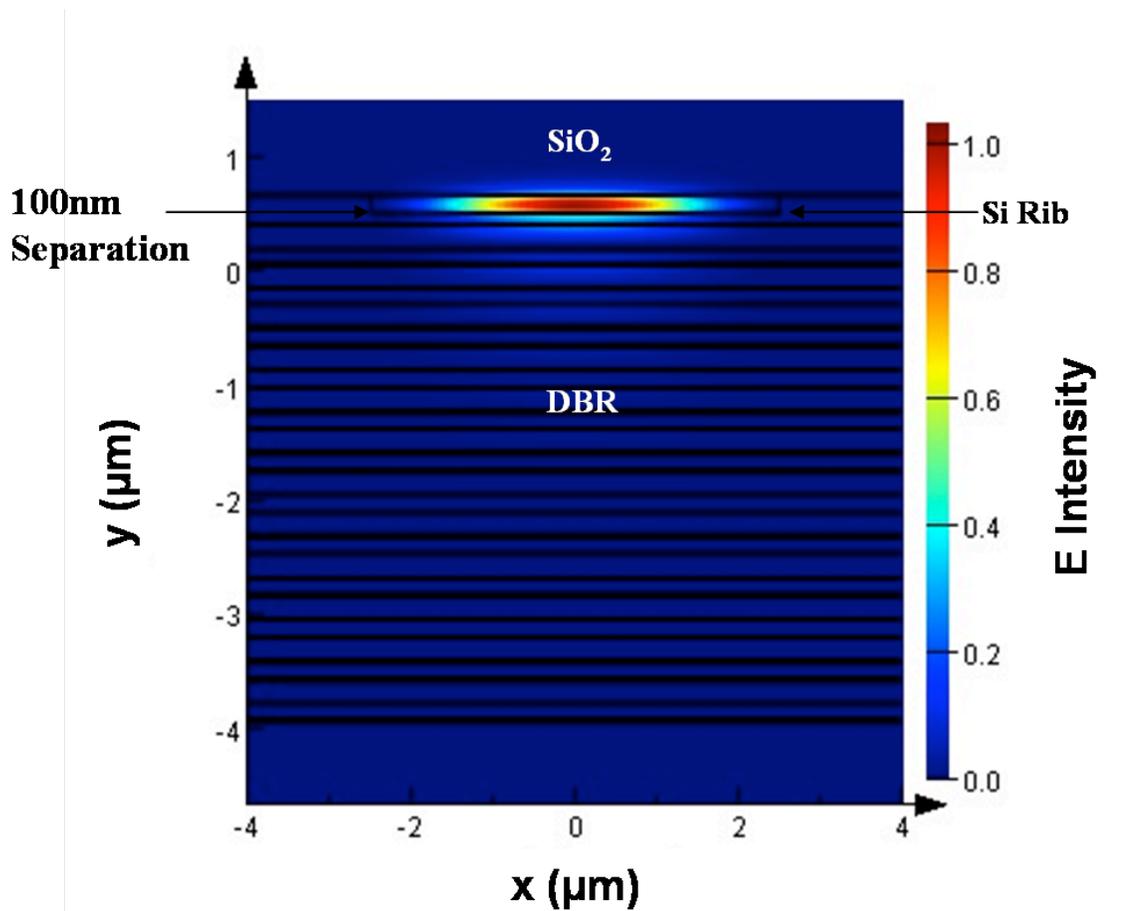


Fig 5-8: 2-d analysis of the stack without any bending. The lower boundary of the DBR is a PML, which gives us an estimate of the losses due to non-ideal reflection. The $5\mu\text{m}$ wide Si rib ensures that the mode doesn't interact with the edges. Also, as the waveguide is bent in subsequent calculations, the obtained modal solutions are very similar to actual disk modes, thereby giving us a cylindrically symmetric technique to simulate the micro-disks.

by the software through standard numerical methods for waveguide bending⁷¹. Even though this technique clearly does not take into account the resonance conditions within the micro-disk, it gives a reasonable approximation for the effective index of the mode *and its propagation loss* which is the most important parameter that we are trying to

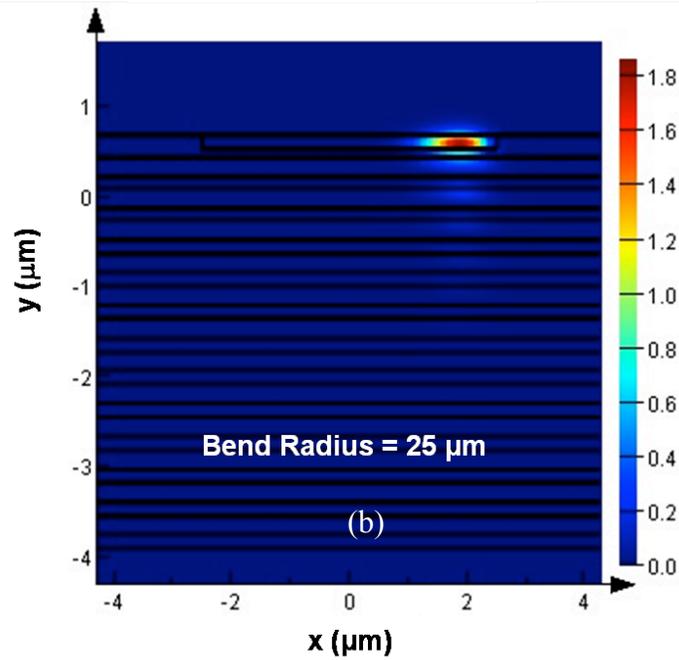
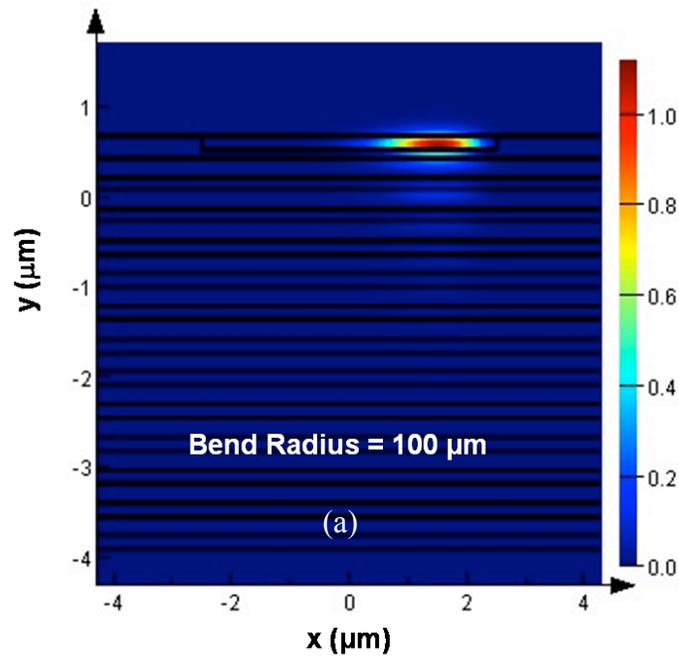


Fig 5-9: 2-d simulations of the Si slab with added curvature. The mode size shrinks to 1 – 2 μm s and is confined close to the outer edge of the disk as expected. The boundaries of the structure are PMLs which gives us estimates for the total bending and substrate losses in the disks

extract from this simulation. Hence this method represents a very efficient way of emulating a full 3-d simulation of the structure by automatically incorporating cylindrical symmetry into the calculation. Fig 5-9 shows the modes obtained for 100 μm and 25 μm radii bends. In agreement with disk resonator theory (see Appendix B) the simulation correctly calculates the mode to be confined within a few microns of the edge (here, since the curvature rotates to the left, the mode is confined to the right corner).

In these simulations the fill factor in the gain layers is found by integrating the Poynting vector within the two GaAsSb layers and normalizing to the total integral across the cross-section of the mode. The fill factor for all the different bend radii was found to be $\sim 5\%$. The losses are estimated by using Perfectly Matched Layers (PML) at the boundaries. The light incident on them is completely absorbed and it can be integrated over time to estimate the loss coefficient for the disk mode. Fig 5-10 shows the losses calculated for various radii used in the simulations. For extremely large disks ($\sim 1000\mu\text{m}$ radius) the loss is dominated by the non-ideal reflection of the DBR and is estimated to be $\sim 4.5/\text{cm}$. Then as the disk gets progressively smaller the bending losses start increasing exponentially and finally for a bend radius $< 25\mu\text{m}$ s they become large enough to render the hybrid laser impractical.

From the above calculations, a useful relationship can be found between the amount of gain required from the GaAsSb layers to achieve lasing and the size of the micro-disk. Since the fill factor in the gain medium is $\sim 5\%$, the total gain needed from the GaAsb is 20 times the total loss in the disk. Therefore from Fig 5-10, for an ideal disk (with infinite Q) that has no extra losses due to parasitic scattering from surface rough-

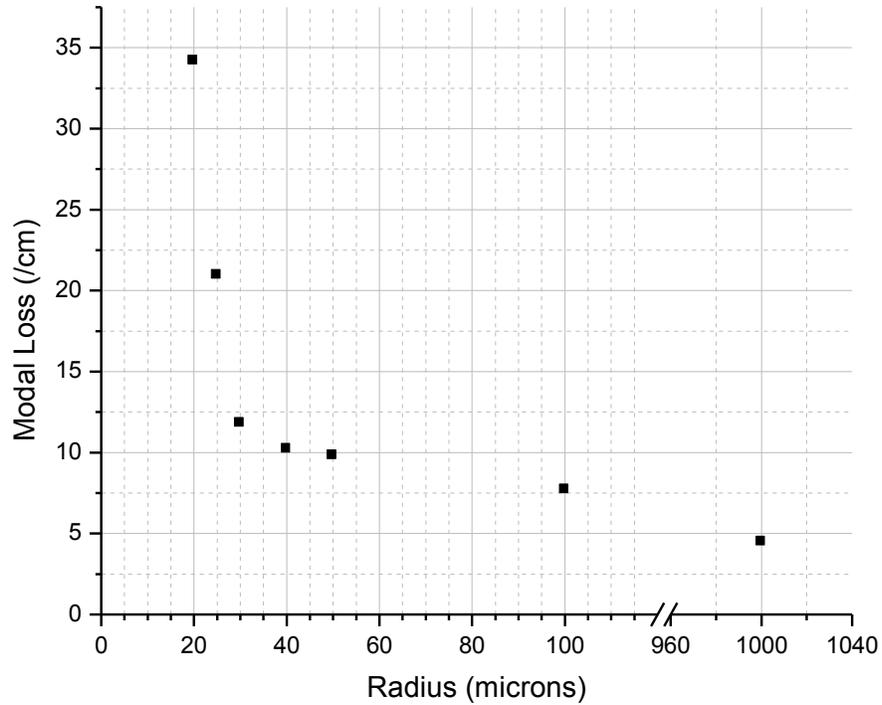


Fig 5-10: Finite Element analysis based estimate of the total losses in ideal Si micro-disks (with no scattering loss) 100nm away from the DBR stack. The major components are the losses through the DBR (which dominate at the large radii) and the bending losses (which dominate for the smaller disks)

ness, the required gain would be 700/cm as the radius reduces to $\sim 20\mu\text{m}$. Even if the disk is non-ideal and has an additional loss of say 15/cm (corresponding to a $Q < 10^4$ which is 2 orders of magnitude less than the state of the art mentioned at the beginning of this chapter), the total loss would be $\sim 50/\text{cm}$ and hence the net required gain would still only be of the order of 1000/cm. This level of gain is readily achievable in most direct-gap semiconductors⁶⁵ and should be reached (@1.55 μm) for carrier injection of the order of $5 \times 10^{18} \text{cm}^{-3}$. *This analysis therefore establishes the feasibility of the evanescent laser*

design as long as the radius of the disk is greater than $20\mu\text{m}$ and the scattering loss in the disk is $<20/\text{cm}$. This estimated lower bound for the disk radius and the corresponding gain requirement also gives us a baseline on which to develop the initial experiments described in the next section.

A crucial point that must be made is that this estimate for the minimum practical disk radius *would be roughly valid even if the separation between the disk and the DBR were changed from the 100nm taken in the simulation.* This is because, as the separation is changed, the total losses and the gain fill factor would scale equally because they are both proportional to the total overlap that the Si mode has with the III-V.

5.3 Towards experimental demonstration

This section briefly describes the steps that were taken towards fabricating and characterizing a hybrid Si/III-V laser. Unfortunately, the experiments didn't lead to fruition and therefore the reasons for not getting lasing are looked at and possible solutions are proposed.

5.3.1 Fabrication

The processing of the device has two important parts: the growth of the DBR stack and the fabrication of the micro-disk. The III-V stack was deposited on an InP wafer using MBE at the University of California, Santa Barbara. The micro-disk fabrication is done using dry etching and extensive experiments are done to characterize its quality factor. Fig 5-11 shows the top view schematic of the structures that are made. The waveguide is included in the design in order to measure the Q of the micro-disk (the

next section describes the reason for this layout and the Q measurement process in much more detail). The tapers are introduced so that the waveguide is single mode in the region close to the disk while being wide enough at the two ends to couple light in and out.

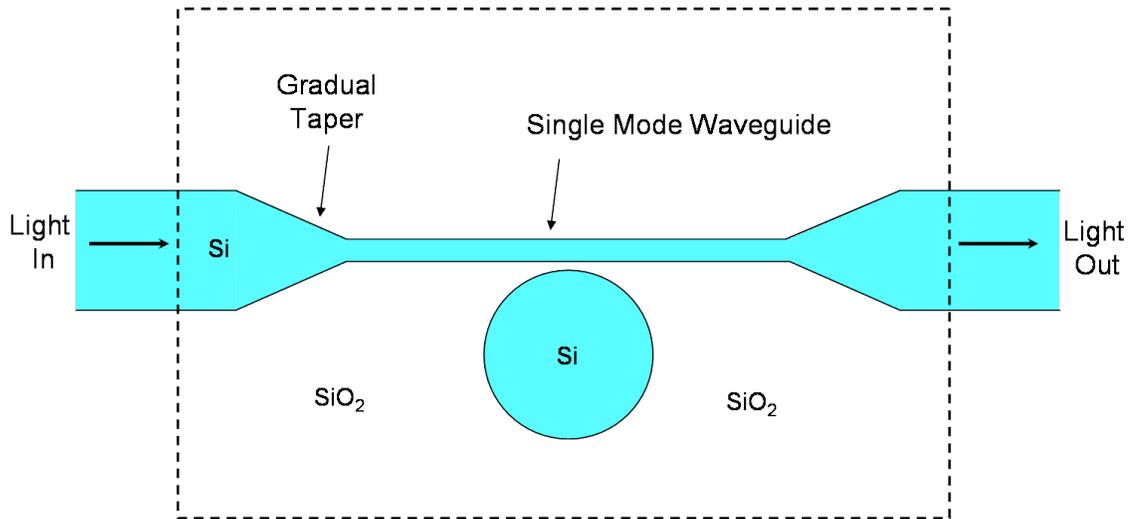


Fig 5-11: Top view schematic of the fabricated structures. The single mode waveguide (500nm width) is put next to the disk in order to measure its Q. The ends of the waveguide are tapered to larger widths (2.5 μ m) to ease the input and output coupling of light. The wider portions of the waveguide extend up to a distance of a few mm. Disks of various radii were fabricated

In the context of the fabrication it must be noted from Fig 5-11 that the device layout has a very long structure (~5 mm) in the form of the waveguide alongside a feature that is of the order of ~100nms (the spacing between the disk and the waveguide) Due to this, both e-beam and photo-lithography are used during the fabrication. The wider (and longer) parts of the waveguide are defined first using photo-lithography with high resolution MiR-703 resist and chlorine based reactive ion etching ($\text{BCl}_3 + \text{Cl}_2 + \text{Ar}$) that etches the Silicon all the way to the buried oxide. The remainder of the device (region

enclosed by the dotted rectangle in Fig 5-11) including the disk and the single mode section of the waveguide, is defined using e-beam lithography and the same dry etching as above. The alignment in the two steps is done using the same marks so that the waveguide is stitched with an accuracy better than 50nm.

As opposed to the gold pads that are typically used, the alignment marks here are in the form of square holes in the Si (etched down to the BOX) that are made *during the first photo-lithography step* itself, thereby automatically aligning them to the waveguide. Then, to improve the visibility of the marks inside the e-beam writer, the exposed buried oxide in the holes is etched a further 1 μm using BOE. This improves the contrast between the top Si and the buried oxide so as to make the alignment much more accurate during the e-beam step. Fig 5-12 shows an SEM close-up of a finished structure with a micro-disk fabricated next to a waveguide.

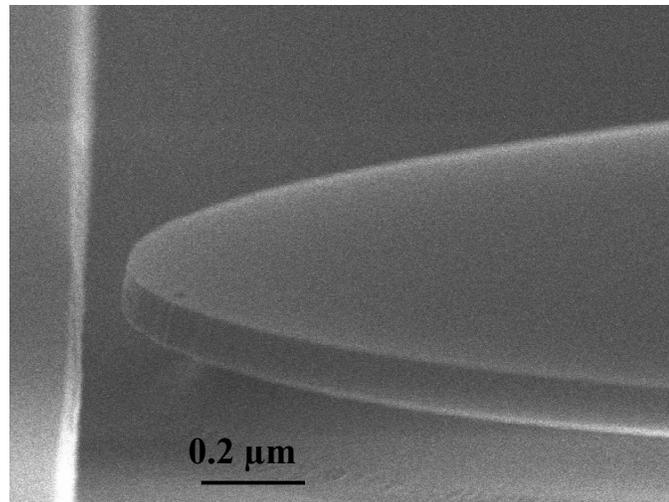


Fig 5-12: Close-up SEM image of the separation of a micro-disk and the coupling waveguide. The structure was made using a combination of e-beam & photo-lithography and dry etching. The undercut in the buried oxide is created by a final wet etch in HF.

After the lithography steps are completed, the waveguides are cleaved at both ends leaving a segment that is a few mm long.

5.3.2 *Q measurement*

The disks are fabricated in close proximity with a waveguide so that a small fraction of the power can couple between the two and this cross-coupling gives us a method to estimate the Q of the disk. From standard resonator theory^{65,72}, it is well established that the transmission spectrum of the waveguide will have dips at points corresponding to the resonance wavelengths of the various modes in the disk. The output spectrum can be converted to the Q of a particular mode using the following expression:

$$Q \approx \frac{\lambda_0}{\Delta\lambda} \quad (5.8)$$

where, λ_0 is the resonance wavelength of the mode in question and $\Delta\lambda$ is the full-width half-maxima of the transmission dip seen at λ_0 . In the context of the laser, we are obviously interested in the values of λ_0 in the 1550nm window.

The Q of the disk depends on the total optical loss, which in turn is comprised of two components: the intrinsic scattering losses in the disk and the coupling of light to the waveguide. Due to the fact that the coupling losses are included in the estimate, the value given by eqn 5.8 represents the *loaded* Q of the resonator as opposed to the *intrinsic* Q , which measures only the inherent loss.

Applying scattering matrix theory⁷², it is found that *at resonance*, the ratio of the output and the input power of the waveguide is:

$$\frac{E_{Out}}{E_{In}} = \frac{C - A}{1 - C A} \quad (5.9)$$

where, C is such that $\sqrt{(1-C^2)}$ represents the fraction of light coupled from the disk to the waveguide and vice versa (the coupling is assumed to be symmetric) and A is the relative amplitude of electric field left after one round-trip (i.e., $\sqrt{(1-A^2)}$ is the fractional round-trip loss inside the disk). Ideally if $C = A$, i.e., if *critical coupling* is achieved, the output falls to 0 and we get the sharpest dip in the spectrum.

The fraction coupled, $k = \sqrt{(1-C^2)}$, strongly depends on the width of the waveguide, the diameter of the disk and the spacing between the two^{72,73,74}. These parameters should be carefully chosen so as to get a coupling strength that is sufficient to see a noticeable dip in the transmission spectrum (given by eqn 5.9), but is not large enough to dominate the total losses in the disk. Appendix C describes the method used to estimate the value of k so that an appropriate spacing can be chosen. In general it is seen that for disks with radii ranging from 10-20 microns, 100nm spacing gives coupling of the order of 1% to a single mode waveguide, which is an ideal amount for this experiment. The waveguide is chosen to be single mode (by making its width ~500nm) for a couple of reasons: firstly the narrow width ensures that the mode inside is more exposed and interacts more readily with the disk. Additionally, the presence of a multi-mode waveguide can needlessly add to the number of resonant dips seen in the transmission spectrum due to parasitic coupling to the higher order modes.

Fig 5-13 shows a schematic of the setup used for the Q measurements. The transmission spectrum is obtained by using a broadband source (a Superluminescent LED

centered around $1.55\mu\text{m}$) at the input of the waveguide and measuring the output with an Optical Spectrum Analyzer (OSA). The OSA was an Ando 6319 with a resolution of 50pm. Lensed optical fibers ($2\mu\text{m}$ beam spot) mounted on micro-positioners are used at the input and output to focus and collect light respectively.

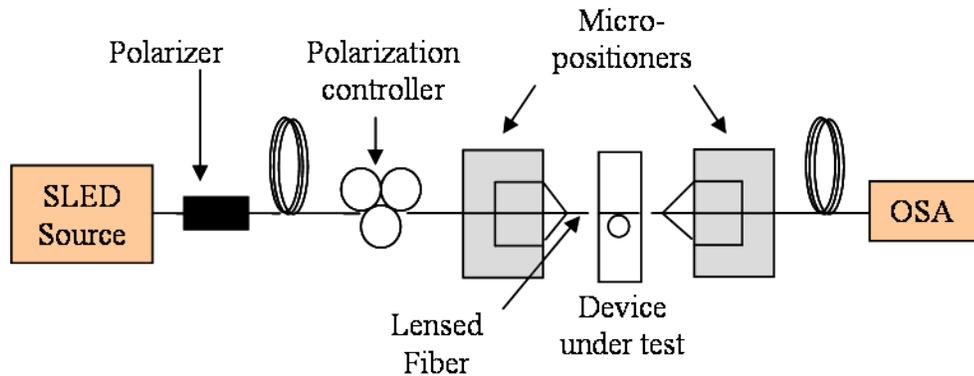


Fig 5-13: Schematic of the Q measurement setup. Broadband light from a SLED is focused into one facet of the cleaved waveguide using a lensed fiber. An identical fiber collects the output light and feeds it into an OSA where the spectrum is analyzed.

Fig 5-14 shows a typical resonant dip obtained for a $20\mu\text{m}$ diameter disk 100 nm away from a single mode waveguide. We can see that the loaded Q is $\sim 10,500$, which corresponds to a loss of $\sim 10.5/\text{cm}$. For a $20\mu\text{m}$ diameter disk that translates into a round-trip loss of $\sim 6\%$. As mentioned before, for the geometry used in this experiment, $\sim 1\%$ of the light gets coupled from the disk to the waveguide and vice versa. Therefore the total loss in the disk is clearly not being overly affected by the coupling to the waveguide and hence the intrinsic Q is also of the same order as the loaded Q.

Qs in the range of 1 to 2×10^4 were typical for disks with diameters between 20 and $60\mu\text{m}$. The main factor limiting the Q is the quality of the dry etch used to define the

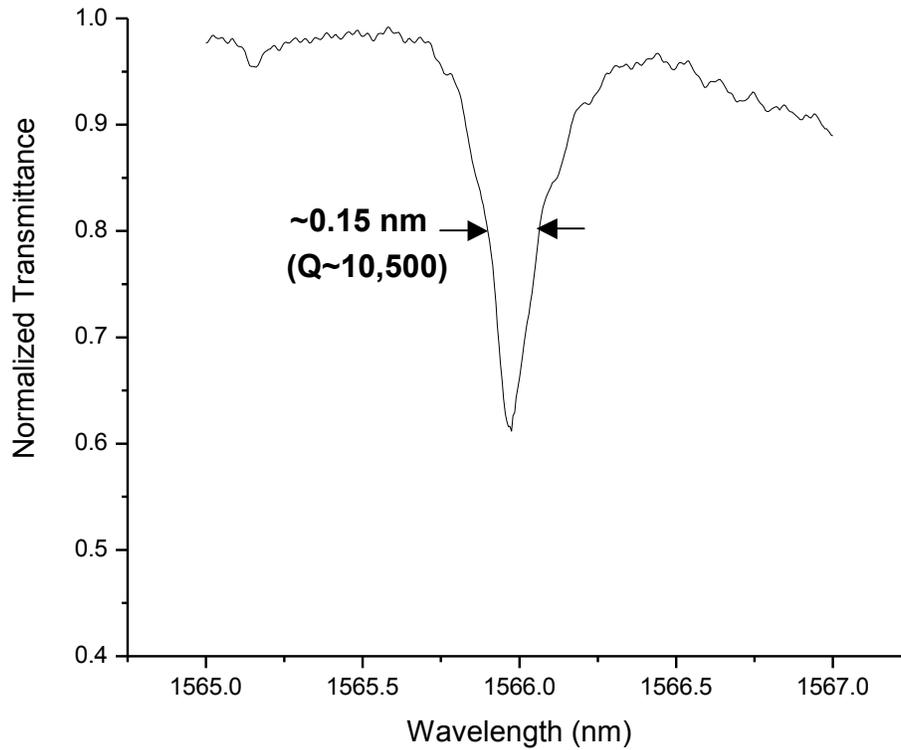


Fig 5-14: Typical transmission spectrum obtained for a 20µm diameter disk 100nm away from the waveguide. The estimated loaded Q is ~10,500.

disks. Fine tuning the RF power used in the RIE chamber can improve surface roughness and give much lower losses. But enhancing the Q was not given high priority because as was discussed in the finite element analysis in section 5.2.2, the losses are already low enough in the context of achieving lasing in the overall structure. Just to reiterate what was said earlier, a 20µm radius with a Q of 10,500 has total losses (through intrinsic scattering, coupling to the waveguide and coupling into the III-V) of ~45/cm, which combined with the 5% fill factor in the gain layers of the DBR, translates to a very reasonable required gain of 900/cm from the GaAsSb alloy.

The next section briefly deals with the practical aspects of demonstrating lasing in the fabricated disks and draws conclusions in terms of the requirements for an efficient III-V/Si hybrid laser.

5.3.3 Lasing requirements and measurement

Once the fabrication is complete, the last step in the experiment is to bring the III-V in close proximity to the Silicon and pump it in order to get lasing. Ideally some form of wafer bonding^{64,75} should be used to join the III-V and the Silicon, but for a proof of concept experiment the two can be brought together using mechanical force exerted on the III-V. This was verified in a simple test where weights were used to apply pressure on a 5mm x 5mm piece of the DBR stack and the separation was monitored through interferometric estimates.

As mentioned in section 5.1.3, the III-V can be pumped electrically or for ease of demonstration, optically, using a laser with a wavelength $>1.1\mu\text{m}$. The electrical pumping will obviously require a detailed study of the contact metallurgy for this alloy system. For the optical pumping a laser in the $1.3\mu\text{m}$ window would be ideal since several high power options are available. CW and pulsed lasers using Nd:YVO₄ (neodymium vanadate) crystals or a lesser transition in the more widely used Nd:YAG crystals can be employed. Assuming a 1ns carrier lifetime in the III-V, normally incident light of the order of $10\text{KW}/\text{cm}^2$ would be needed to achieve transparency. Finally, the waveguide etched close to the disk for the Q estimates can be used to collect the output light from the hybrid laser.

The outline described above should be successful, but similar tests that we conducted didn't exhibit lasing. This was because in our case, tests on the DBR stack grown using MBE gave very poor photo-luminescence efficiency @1.55 μm (<0.01%) and further epitaxial runs weren't performed due to extraneous circumstances. The use of the atypical Antimony based alloy system was the main reason for the lack of options to carry out the growth. It should be noted that the emission efficiency is too low to get lasing from even an ideal micro-disk with extremely high Q, because the bending and reflection losses wouldn't be overcome.

Since photo-spectrometric measurements confirmed that the thicknesses in the DBR were according to the specifications, the lack of gain available from the III-V *was the primary reason for the failure of the lasing experiments.*

5.3 Conclusions

In this chapter we have theoretically established the feasibility of making a hybrid Si/III-V laser @1.55 μm based on evanescent coupling. By introducing a III-V stack that provides gain and acts as a DBR, lasing can be expected in micro-disks with mediocre quality factors. The device design relaxes some of the constraints on the introduction of a III-V alloy into a Silicon foundry and takes a step towards a CMOS compatible laser solution on Si that can be electrically pumped.

The computational modeling based on finite element analysis gives us a framework within which the device parameters can be finalized. Unfortunately, we were not able to exhibit lasing in the proof-of-concept tests that were conducted, mainly due to

poor luminescence efficiency in the DBR stack. But it must be stressed that the overall gain requirements are not that stringent and this should therefore not be considered an insurmountable obstacle in the path toward demonstrating the hybrid lasers proposed here. Further exploration of the factors affecting gain in epitaxial Antimonide films and a thorough assessment of the other material options available should help greatly in resolving the challenges that were faced here.

Chapter 6: Conclusions

In this dissertation we have looked at novel Silicon based devices for both detection and emission of light in the optical communication window, that have the potential to be fabricated with CMOS compatible processes. The need for photonic components on Si has risen due to the fact that optical interconnects are seen as a viable alternative to electrical connections in standard integrated circuits. Electrical interconnects are now facing problems in efficiency because of increased power consumption, higher crosstalk and longer latency.

The experiments reported in the first part of this thesis explored the feasibility of using low temperature Ge epitaxy on Silicon for the purposes of photo-detection. Due to the 4% lattice mismatch between Si and Ge, the biggest problem that we faced in our work was the reduction of dislocations in Ge layers grown directly on Si. In addition, we focused our attention on films that could be processed at lower temperatures ($<450^{\circ}\text{C}$) so that the devices proposed later in the dissertation can be fabricated in a standard Si foundry without significantly altering the process flow. Hence we primarily used molecular beam epitaxy at 370°C to deposit the Germanium (p-Ge on n-Si substrates). The resultant films were characterized through standard structural analysis techniques (X-ray diffraction and Transmission Electron Microscopy) and also through an indirect measurement of electron diffusion length in the Germanium. The diffusion length, which was estimated from the responsivity of simple normal-incidence detectors, was used to

gauge quality because it directly correlates to the density of dislocations in the Ge film. It was seen that the substrate surface preparation and Ge growth rate are critical in determining the crystalline quality of the epitaxial Ge. By modifying the deposition rate and adopting various Si substrate surface treatments for the removal of hydrogen and native oxide, we obtained diffusion lengths of up to 50nm. These lengths compare favorably with previous studies that used alternative deposition techniques.

The next part of the dissertation discussed the Photo Heterojunction FET; a novel detector based on the Ge/Si films described above. The design of the device consisted of an FET structure where the gate on the Silicon was replaced by a Germanium island. Infra-red light incident on the Ge creates charge separation due to the unique band alignment between Silicon and Germanium and this in turn modulates the conductance of the Si channel, thereby emulating simple transistor behavior.

The great advantage of the device lies in the fact that there is no contact on the Ge which in turn presents unique scaling opportunities, since the gate length could easily be reduced to state of the art MOSFET dimensions. The reduction in capacitance that would accompany the scaling would greatly improve the sensitivity of a receiver that employs the device. The lower optical absorption cross-section of a smaller gate can be countered by embedding the detector in a cavity. The design of an efficient cavity is in fact a major component of the future work that must be done to fully realize the potential of the proposed device. Finally it must also be mentioned that there are no roadblocks from a fabrication standpoint, since Germanium has already been widely adopted by standard CMOS foundries to provide strain to the Si channel

In the proof-of-concept experiments that were conducted, detectors with channel lengths of the order of a micron, exhibited very high sensitivity to CW light. The Ge gate increased the conductance of the Si channel by a factor of 25 in spite of absorbing only $\sim 100\text{nW}$ of the incident $1.55\mu\text{m}$ light. The high speed performance was also encouraging and rise times as low as 40ps were observed. The main caveat was that the fall times were of the order of 10s of ns and even though that represents a significant problem, it is one that can be resolved with innovative device and circuit designs. Overall, the proposed detector can provide vast improvement in receiver sensitivity with practically limitless scaling possibilities.

The final part of the thesis dealt with the design of a novel III-V/Si hybrid laser in the optical communications window based on evanescent coupling. The idea stemmed from recent demonstrations of high Q Si micro-disks with losses of the order of $0.1/\text{cm}$. The low loss implied that these disks could be used to fabricate lasers that require very little gain to reach threshold. But even this minimal amount of gain has to be extracted from an external source, since Silicon is famously inefficient at light emission. The device proposed in chapter 5 tackled this problem by using proximity gain wherein a III-V stack is brought close enough to a Silicon disk so as to overlap with the cavity mode and provide gain. Losses through the high index gain region were avoided by a unique design in which the III-V gain stack also doubled as a DBR for the evanescent tails of the mode in the Si micro-disk. The choice of Antimonide based alloys (lattice matched to InP) ensured that the DBR provided gain at $1.55\mu\text{m}$ and that there was a high enough index contrast between its constituent layers to not require overly thick epitaxial growth.

All things considered, this design greatly eases the restrictions that are normally associated with the incorporation of III-V material into a CMOS foundry.

From a modeling perspective, transfer matrix formulation and finite element analysis predicted exceptional reflection characteristics for the designed DBR. In fact the dominant loss mechanism was found to be bending losses in the micro-cavity which occur due to the lack of lateral confinement in the stack. From the numerical analysis, it was estimated that in order to keep the required gain in the III-V below 1000/cm, the micro-disk radius should be at least 20 μ m.

The above calculations provided us with guidelines to pick the design parameters for a demonstration of the device, but unfortunately proof-of-concept experiments to that end were unsuccessful. Disks with quality factors of about 10^4 were fabricated but didn't exhibit lasing, primarily due to insufficient gain in the III-V layers. Because of the unconventional nature of the Antimonide material system, repeated iterations of the epitaxial growth of the DBR stack were precluded and lasing could not be shown. But in any case, our research into this device does indicate that the gain requirements of the system are well within reach and access to high quality epitaxial material should pave the way towards a fully functional III-V/Si hybrid evanescent laser.

In conclusion, this dissertation has contributed novel ideas towards the fabrication of efficient photo-detectors and lasers on CMOS chips for Silicon photonics applications. These devices have unique advantages that can prove to be extremely useful as optical integration becomes increasingly relevant in the electronics industry.

Appendix A: Estimating the band-gap and refractive index of AlGaAsSb

The refractive index of a semiconductor can be obtained from a simplified interband-transition model that ignores excitonic effects⁶⁸. The expression for the index at a frequency corresponding to photon energy E , is as follows:

$$n(E)^2 = A_0^* \left[f(\chi_0) + \frac{1}{2} \left(\frac{E_0}{E_0 + \Delta_0} \right)^{\frac{3}{2}} f(\chi_{so}) \right] + B^* \quad (\text{A.1})$$

where,

$$f(\chi_0) = \chi_0^{-2} \left[2 - (1 + \chi_0)^{\frac{1}{2}} - (1 - \chi_0)^{\frac{1}{2}} \right]$$

$$f(\chi_{so}) = \chi_{so}^{-2} \left[2 - (1 + \chi_0)^{\frac{1}{2}} - (1 - \chi_{so})^{\frac{1}{2}} \right]$$

$$\chi_0 = \frac{E}{E_0} \quad \chi_{so} = \frac{E}{(E_0 + \Delta_0)}$$

Here, E_0 is the band-gap, Δ_0 is the spin-orbit splitting, A_0^* is the strength of the E_0 & $E_0 + \Delta_0$ free electron-hole transitions and B^* is the strength of the transitions from the higher gaps. For a quaternary alloy each of these 4 parameters can be derived from its constituent ternary semiconductors. The ternary alloy parameters can in turn be calculated from the values for their respective binary components.

Therefore for $\text{Al}_x\text{Ga}_{1-x}\text{As}_y\text{Sb}_{1-y}$ this will take the following form:

$$Q(x, y) = \frac{\left\{ \begin{aligned} &x(1-x) \left[yT_{\text{AlGaAs}}(x) + (1-y)T_{\text{AlGaSb}}(x) \right] \\ &+ y(1-y) \left[xT_{\text{AlAsSb}}(y) + (1-x)T_{\text{GaAsSb}}(y) \right] \end{aligned} \right\}}{x(1-x) + y(1-y)}$$

$$T_{\text{AlGaAs}}(x) = xB_{\text{AlAs}} + (1-x)B_{\text{GaAs}} + C_{(\text{Al,Ga})\text{As}}(x)(1-x)$$

$$T_{\text{AlGaSb}}(x) = xB_{\text{AlSb}} + (1-x)B_{\text{GaSb}} + C_{(\text{Al,Ga})\text{Sb}}(x)(1-x) \quad (\text{A.2})$$

$$T_{\text{AlAsSb}}(y) = yB_{\text{AlAs}} + (1-y)B_{\text{AlSb}} + C_{\text{Al}(\text{As,Sb})}(y)(1-y)$$

$$T_{\text{GaAsSb}}(y) = yB_{\text{GaAs}} + (1-y)B_{\text{GaSb}} + C_{\text{Ga}(\text{As,Sb})}(y)(1-y)$$

where, the T 's, B 's & Q represent the parameter being calculated (E_0 , Δ_0 , A_0^* or B^*) and the C 's are the 'Bowling' parameters that quantify the deviation from an idealized interpretation of Vegard's law⁶⁸. We therefore start with the values for the simple alloys AlAs, AlSb, GaSb and GaAs and finally arrive at the value for $\text{Al}_x\text{Ga}_{1-x}\text{As}_y\text{Sb}_{1-y}$ by plugging the composition (x and y) into the equations in A.2. For epitaxy on InP, the following lattice matching expression gives the value of y for a given x :

$$y = (0.227 + 0.04x)/(0.443 + 0.033x) \quad (\text{A.3})$$

The values for E_0 , Δ_0 , A_0^* or B^* for the four binary alloys are listed in refs 68 and 76.

Therefore to conclude, eqns A.2 and A.3 are used to calculate the band-gap, spin-orbit split off and transition strengths for $\text{Al}_x\text{Ga}_{1-x}\text{As}_y\text{Sb}_{1-y}$ from its 4 constituent binary alloys. These values are then plugged into eqn A.1 to find the refractive index of the alloy at any wavelength.

Appendix B: Analytic solutions for the modes of a disk resonator

In order to find the modes of a micro-disk we first start with the cylindrical form of the time-independent Maxwell's wave equation⁶⁶:

$$\left(\frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} + \left(\frac{\omega}{c} \right)^2 n^2(\mathbf{r}) \right) F(\mathbf{r}) = 0 \quad (\text{B.1})$$

where, $n(\mathbf{r})$ represents the refractive index inside the disk. As mentioned in section 5.1.2, the high index contrast between Silicon and the surrounding air/oxide makes two Si slab modes dominant (TE and TM). F can therefore be expressed as $F_z = W(\rho, \phi) Z(z)$ with the 'z' subscript indicating the transverse nature of the electric or magnetic field (F_z corresponds to H_z for TE modes and E_z for TM modes). This reduces the above equation to:

$$\frac{1}{W} \left(\frac{\partial^2 W}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial W}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial W^2}{\partial \phi^2} \right) + \frac{1}{Z} \frac{d^2 Z}{dz^2} + k_0^2 n^2(\mathbf{r}) = 0 \quad (\text{B.2})$$

where, k_0 is simply ω/c . Applying separation of variables, we get:

$$\frac{1}{W} \left(\frac{\partial^2 W}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial W}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial W^2}{\partial \phi^2} \right) + k_0^2 n_{eff}^2 = 0 \quad (\text{B.3(a)})$$

$$\frac{1}{Z} \frac{d^2 Z}{dz^2} + k_0^2 (n_{Si}^2 - n_{eff}^2) = 0 \quad (\text{B.3(b)})$$

where, $k_0^2 n_{eff}^2$ is the constant that both equations are equal to (n_{eff} is the slab index in the Silicon). Eqn B.3(b) is simply the wave equation for the slab. B.3(a) can be solved by an

additional separation of variables⁶⁶ given by $W(\rho, \phi) = \psi(\rho) \Omega(\phi)$. The equations for ψ and Ω take the form:

$$\frac{\partial^2 \psi}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial \psi}{\partial \rho} + \left(k_0^2 n_{eff}^2(\rho) - \frac{m^2}{\rho^2} \right) \psi = 0 \quad (\text{B.4(a)})$$

$$\frac{\partial^2 \Omega}{\partial \phi^2} + m^2 \Omega = 0 \quad (\text{B.4(b)})$$

where, m is the angular mode number. The equations in B.3(b) and B.4 must be solved self-consistently to extract the wavelength (i.e., k_0) for a given mode of order m . B.4(a) is a standard radial wave equation whose solution ψ can be approximated by a Bessel function for $\rho < R$ and by a decaying exponential for $\rho > R$. Therefore:

$$\varphi(\rho) \sim \begin{cases} J_m(k_0 n_{eff} \rho) & \rho \leq R \\ J_m(k_0 n_{eff} R) \exp(-\alpha(\rho - R)) & \rho > R \end{cases} \quad (\text{B.5})$$

where, $\alpha = k_0 (n_{Si}^2 - n_{eff}^2)^{1/2}$. Using B.5 in B.4 and applying the appropriate boundary equations we get the following transcendental equation that can be solved to get k_0 for a given m :

$$k_0 n_{eff}(k_0) J_{m+1}(k_0 n_{eff}(k_0) R) = \left(\frac{m}{R} + \eta \alpha \right) J_m(k_0 n_{eff}(k_0) R) \quad (\text{B.6})$$

where, η is n_{eff}^2 / n_{Si}^2 for TE modes and 1 for TM modes. The last two equations *give us the radial form of the modal field and the wavelength at which each resonance occurs.*

An important conclusion that must be made from the functional form given in eqn B.5 is that the modal profile (for the first order radial modes) will peak close to the edge of the disk (due to the nature of the Bessel function) and *therefore the light is primarily confined at the rim of the micro-disk.*

Appendix C: Estimating disk-waveguide coupling strength analytically

The coupling between a micro-disk and a waveguide that are in close proximity can be calculated from coupled mode theory for two waveguides⁷⁷. The z dependence of the fields in the disk and waveguide is given by:

$$\begin{aligned} \pm \frac{dE_{Disk}}{dz} &= i\kappa_{dd} E_{Disk} + i\kappa_{dw} E_{WG} e^{i(\beta_w - \beta_d)z} \\ \pm \frac{dE_{WG}}{dz} &= i\kappa_{ww} E_{WG} + i\kappa_{wd} E_{Disk} e^{i(\beta_d - \beta_w)z} \end{aligned} \quad (C.1)$$

where, z is the coordinate along the direction of propagation, κ_{dd} and κ_{ww} are the self-coupling coefficients, κ_{dw} and κ_{wd} are the cross-coupling coefficients and β_w & β_d are the propagation constants for the modes in the waveguide and disk respectively. A few observations must be made before the equations in C.1 can be integrated:

1. The spacing between the waveguide and the disk *changes with z* , due to the curvature of the disk. The dependence of the spacing on z can be approximated by a parabolic function⁷³ of the form:

$$s(z) = s_0 + \frac{z^2}{2R} \quad (C.2)$$

where, s_0 is the minimum distance between the waveguide and the disk and R is the disk radius. The effect of this is that the coupling coefficients (which are standard overlap integrals⁷⁷) also become dependant on z :

$$\kappa_{ab}[s(z)] = \frac{\epsilon_0 \omega}{4} \int_{-\infty}^{\infty} e_a^*(x) \Delta \epsilon_b e_b(x) dx \quad (C.3)$$

where, x is the coordinate perpendicular to the direction of propagation and e_a & e_b are power normalized field profiles (the a and b subscripts represent the disk or the waveguide depending on which coupling coefficient is being calculated). The positions of e_a and e_b in terms of x will change with z , thereby changing the overlap integral.

2. The propagation constant for the disk is calculated for a mode (of order m) by satisfying the resonance condition at the rim of the disk:

$$2 n_{\text{mode}} \pi R = m \lambda \quad \text{i.e.,} \quad \beta_d R = m \quad (\text{C.4})$$

Obviously the modes that couple most strongly are the ones that have a value of β close to the propagation constant in the waveguide (β_w).

3. The profile of the disk mode (e_a or e_b in eqn C.3) in terms of x is a Bessel function inside the disk and a decaying exponential outside (as seen in Appendix B).

Now that all the necessary parameters are in place, the overall fraction of light coupled from the waveguide to the disk, at resonance, can be estimated. The propagation constant (β_w) of the waveguide is first calculated at the micro-disk resonance wavelength of interest (λ_0). This can be done using any commercial 2-d mode solver. β_d and all the κ 's at each z are then calculated using eqns C.4 and C.3 respectively, for disk modes that have roughly similar propagation constants to the mode in the waveguide. These values are then plugged into the coupled equations in C.1, which are in turn integrated numerically (in our work Matlab® was used for this purpose). The ratio of the final and initial value of $|E_{WG}|^2$ yields the fraction of power left in the waveguide thereby giving the amount coupled into the disk.

References

1. E.A. Fitzgerald and L.C. Kimmerling, *MRS Bulletin*, Vol. 23, No.4, p. 39 (1998).
2. J.A. Davis and J.D. Meindl, *Interconnect Technology and Design for Gigascale Integration*, Springer, 2003.
3. D.A.B. Miller, *Optics Letters*, Vol. 14, No. 2, p. 146 (1989).
4. A. Bandyopadhyay and M.J. Deen, *Photodetectors for Optical Fiber Communications*, Chapter 5, Photodetectors and Fiber Optics, ed. H.S. Nalwa, Academic Press, New York (2001).
5. G.P. Watson, E.A. Fitzgerald, Y.H. Xie and D. Monroe, *Journal of Applied Physics*, Vol. 75, No. 1, p. 263 (1994).
6. S. Famà, L. Colace, G. Masini, G. Assanto, and H-C. Luan, *Applied Physics Letters*, Vol. 81, No. 4, p. 586 (2002).
7. H-C. Luan, K.Wada, L.C. Kimmerling, G. Masini, L.Colace, and G. Assanto, *Optical Materials*, Vol. 17, p. 71 (2001).
8. G. Masini, L. Colace, F. Galluzzi and G. Assanto, *Materials Science and Engineering B*, Vol. 69-70, p. 257 (2000).
9. G. Masini, L. Colace and G. Assanto, *Applied Physics Letters*, Vol. 82, No. 15, p. 2524 (2003).
10. G. Masini, L. Colace and G. Assanto, *Optics Materials*, Vol. 17, p. 243 (2001).
11. S. Lombardo, S.U. Campisano, G.N. van denHoven, A. Cacciato and A. Polman, *Applied Physics Letters*, Vol. 63, p. 1942 (1993).

12. M. Fujii, M. Yoshida, Y. Kanzawa, S. Hayashi and K. Yamamoto, *Applied Physics Letters*, No. 71, p. 1198 (1997).
13. P.G. Kik, M.L. Brongersma and A. Polman, *Applied Physics Letters*, Vol. 76, p. 2325 (2000).
14. H.S. Han, S.Y. Seo and H.J. Shin, *Applied Physics Letters*, Vol. 79, p. 4568 (2001).
15. M.L. Brongersma, A. Polman, K.S. Min, T. Tambo and H.A. Atwater, *Applied Physics Letters*, Vol. 72, p. 2577 (1998).
16. F. Iacona, G. Franzo and C. Spinella, *Journal of Applied Physics*, Vol. 87, p. 1295 (2000).
17. L. Pavesi, L.D. Negro, C. Mazzoleni, G. Franzo and F. Priolo, *Nature*, Vol. 408, p. 440 (2000).
18. T. Trupke, J. Zhao, A. Wang, R. Corkish, and M. Green, *Applied Physics Letters*, Vol. 82, p. 2996 (2003).
19. O. Boyraz and B. Jalali, *Optics Express*, Vol. 12, No. 21, p. 5269 (2004).
20. R. Haisheng, R. Jones, A. Liu, O. Cohen, D. Hak, A. Fang and M. Paniccia, *Nature*, Vol. 433, p. 725, (2005).
21. M. Borselli, K. Srinivasan, P.E. Barclay and O. Painter, *Applied Physics Letters*, Vol. 85, p. 3693 (2004).
22. M. Soltani, S. Yegnanarayanan and A. Adibi, *Optics Express*, Vol. 15, No. 8, p. 4694 (2007)
23. J. Singh, *Electronic and Optoelectronic Properties of Semiconductor Structures*, Cambridge University Press, 2003.

24. G. Masini, L. Colace and G. Assanto, Chapter 7 in *Handbook of Thin Film Materials, Vol 4: Semiconductor and Superconductor Thin Films*, ed. H.S. Nalwa, Academic Press, 2002.
25. J. Tersoff, *Physical Review B*, Vol. 43, No. 11, p. 9377 (1991)
26. F. K. LeGoues, M. Copel and R. M. Tromp, *Physical Review B*, Vol. 42, No. 18, p. 11690 (1990)
27. G. Masini, L. Colace, G. Assanto, H-C. Luan, and L.C. Kimmerling, *IEEE Transactions on Electron Devices*, Vol. 48, No. 6, p. 1092 (2001) and references therein.
28. W.G. Oldham and A.G. Milnes, *Solid State Electronics*, Vol. 7, p. 153 (1964)
29. S.M. Sze, *Physics of Semiconductor Devices (2nd edition)*, John Wiley & Sons, 1981
30. J. Weertman and J.R. Weertman, *Elementary Dislocation Theory*, Oxford University Press, 1992.
31. H. Kroemer, T.Y. Liu and P.M. Petroff, *Journal of Crystal Growth*, Vol. 95, p. 96 (1989)
32. P. Bhattacharya, *Semiconductor Optoelectronic Devices*, Prentice Hall, 1994.
33. A. Sakai and T. Tatsumi, *Applied Physics Letters*, Vol. 64, No. 1, p. 52 (1994)
34. J. L. Liu, S. Tong, Y. H. Luo, J. Wan and K. L. Wang, *Applied Physics Letters*, Vol. 79, No. 21, p. 3431 (2001)
35. M. Yamaguchi, *Journal of Materials Research*, Vol. 6, No.2, p. 376 (1991)
36. H.-C. Luan, D. R. Lim, K. K. Lee, K. M. Chen, J. G. Sandland, K. Wada and L. C. Kimmerling, *Applied Physics Letters*, Vol. 75, No. 19, p. 2909 (1999)

37. G.W. Trucks, K. Raghavachari, G.S. Higashi and Y.J. Chabal, *Physical Review Letters*, Vol. 65, p. 504 (1990).
38. V.Le Thanh, *Thin Solid Films*, Vol. 321, p. 98 (1998).
39. D.J. Eaglesham, G.S. Higashi and M. Cerullo, *Applied Physics Letters*, Vol. 59, p. 685 (1991).
40. J. Ihm, M.L. Cohen and D.J. Chadi, *Physical Review B*, Vol. 21, p. 4592, (1980).
41. M. Ohring, *The Materials Science of Thin Films*, Academic Press, San Diego (2002).
42. E. Yablonovitch, *Journal of the Optical Society of America*, Vol. 72, No. 7, p. 899 (1982).
43. A.S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley, New York, 1967.
44. M. Shur, *Physics of Semiconductor Devices*, Prentice Hall, 1990.
45. S. Cristoloveanu and S. Li, *Electrical Characterization of Silicon-on-Insulator Materials and Devices*, Springer, 1995.
46. G. S. Oehrlein, T. D. Bestwick, P. L. Jones, 1 M. A. Jaso and J. L. Lindstrom, *Journal of the Electrochemical Society*, Vol 138, No 5, p. 1443 (1991).
47. S.A. Campbell, *The Science and Engineering of Microelectronic Fabrication*, Oxford University Press, USA, 2001.
48. G.L. McVay and A.R. DuCharme, *Physical Review B*, Vol. 9, No. 2, p. 627 (1974)
49. G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*, Wiley, 2003.
50. D.C. Streit, *Hyperabrupt doping profiles in silicon by molecular beam epitaxy using solid phase epitaxial regrowth*, PhD Thesis, University of California, Los Angeles, 1986.

51. D.Y.C Lie, *Journal of Electronic Materials*, Vol. 27, No. 5, p. 377 (1998)
52. G. L. Olson and J. A. Roth, *Materials Science Reports*, Vol. 3, p. 1 (1988).
53. E. Sackinger, *Broadband Circuits for Optical Fiber Communication*, Wiley, 2005
54. I. Song, *Multi-Gbit/s Transimpedance Amplifier with Integrated Photodetector for Optional Interconnects*, PhD Thesis, Georgia Institute of Technology, 2004.
55. Y.J. Chen and M.D. Plessis, *Microelectronics Journal*, Vol. 37, p. 985 (2006).
56. C. Kromer, G. Sialm, C. Berger, T. Morf, M.L. Schmatz, F. Ellinger, D. Erni, G.L. Bona and H. Jackel, *IEEE Journal of Solid-State Circuits*, Vol. 40, No. 12, p. 2667 (2005).
57. B. Analui, D. Guckenberger, D. Kucharski and A. Narasimha, *IEEE Journal of Solid-State Circuits*, Vol. 41, No. 12, p. 2945 (2006).
58. K. Schneider and H. Zimmermann, *Highly Sensitive Optical Receivers*, Springer, 2006.
59. M. Borselli, T.J. Johnson and O. Painter, *Applied Physics Letters*, Vol. 86, p. 131114 (2006)
60. H. Park, A.W. Fang, S. Kodama and J.E. Bowers, *Optics Express*, Vol. 13, No. 23, p. 9460 (2005).
61. A.W. Fang, H. Park, R. Jones, O. Cohen, M.J. Paniccia and J.E. Bowers, *IEEE Photonics Technology Letters*, Vol. 18, No. 10, p. 1143 (2006).
62. A.W. Fang, H. Park, O. Cohen, R. Jones, M.J. Paniccia and J.E. Bowers, *Optics Express*, Vol. 14, No. 20, p. 9203 (2006)

63. A. Black, A.R. Hawkins, N.M. Margalit, D.I. Babić, A.L. Holmes, Jr., Y.L. Chang, P. Abraham, J.E. Bowers and E.L. Hu, *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 3, No. 3, p. 943 (1997).
64. D. Pasquariello and K. Hjort, *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 8, No. 1, p. 118 (2002).
65. B.E.A. Saleh and M.C. Teich, *Fundamentals of Photonics, 2nd Edition*, John Wiley and Sons, 2007
66. M. Borselli, T.J. Johnson and O. Painter, *Optics Letters*, Vol. 13, No. 5, p. 1515 (2005)
67. M.K. Chin, D.Y. Chu and S.T. Ho, *Journal of Applied Physics*, Vol. 75, No. 7, p. 3302 (1994).
68. S. Adachi, *Optical Constants of Crystalline and Amorphous Semiconductors: Numerical Data and Graphical Information*, Kluwer Academic Publishers, 1999.
69. O. Ostinelli, G. Almuneau and W. Bachtold, *Semiconductor Science and Technology*, Vol. 21, p. 681 (2006).
70. Kramers-Kronig relations in ref 65.
71. G.T. Reed and A.P. Knights, *Silicon Photonics: An Introduction*, Wiley, 2004.
72. R. Grover, *Indium Phosphide Optical Micro-ring Resonators*, PhD Thesis, University of Maryland, College Park, 2003.
73. B.E. Little, S.T. Chu, H.A. Haus, J. Foresi and J.P. Laine, *Journal of Lightwave Technology*, Vol. 15, No. 6, p. 998 (1997).

74. S.C. Hagness, D. Rafizadeh, S.T. Ho and A. Taflove, *Journal of Lightwave Technology*, Vol. 15, No. 11, p. 2154 (1997).
75. A. Ploszl and G. Krauter, *Materials Science and Engineering (Reports)*, Vol. R25, p. 1 (1999).
76. M. Guden and J. Pripek, *Modelling and Simulation in Materials Science and Engineering*, Vol. 4, p. 349 (1996).
77. Jia-Ming Liu, *Photonic Devices*, Cambridge University Press, 2005.