Optical Studies of Bandstructure and Spin-Dependent Processes in Mid-Infrared Semiconductor Materials and Devices

by

Martin Neil Merrick

Submitted for the degree of Doctor of Philosophy

Advanced Technology Institute
School of Electronics and Physical Sciences
University of Surrey
March 2006
© Martin Neil Merrick 2006
Contents

Abstract ........................................................................................................iv
Acknowledgements .......................................................................................vi
Publications .....................................................................................................viii

1 Introduction ...............................................................................................1
  1.1 Optical Characterisation in the Infrared .................................................1
    1.1.1 Background ...................................................................................1
    1.1.2 The Czerny-Turner Grating Spectrometer ......................................4
      1.1.2.1 Principle of Operation .............................................................4
      1.1.2.2 Advantages and Disadvantages .............................................5
      1.1.2.3 Systems Used in this Work .....................................................6
    1.1.3 The Michelson Design FTIR Interferometer .....................................7
      1.1.3.1 Principle of Operation .............................................................7
      1.1.3.2 Advantages and Disadvantages .............................................9
      1.1.3.3 Systems Used in this Work .....................................................10
  1.2 Optically Probing Electron Spin Dynamics ...........................................11
    1.2.1 Background ..................................................................................11
    1.2.2 The FELIX Free Electron Laser System ........................................12
      1.2.2.1 Principle of Operation .............................................................13
      1.2.2.2 Capabilities ............................................................................14
    1.2.3 The Coherent Ti:Sapphire Laser System ........................................16
      1.2.3.1 Principle of Operation .............................................................16
      1.2.3.2 Capabilities ............................................................................18
  1.3 References ...........................................................................................19

2 Optical Characterisation Techniques ........................................................23
  2.1 Introduction .........................................................................................23
  2.2 Photoluminescence ..............................................................................24
    2.2.1 Theoretical Background ..............................................................24
    2.2.2 Practical Implementation ..............................................................31
  2.3 Photomodulation Spectroscopy ...............................................................33
    2.3.1 Theoretical Background ..............................................................33
    2.3.2 Practical Implementation ..............................................................42
  2.4 References ...........................................................................................44

3 First Application of FTIR Spectroscopy to Photomodulation Techniques ....47
  3.1 Introduction .........................................................................................47
    3.1.1 Choosing a Suitable FTIR System ...............................................47
  3.2 Theory ..................................................................................................50
    3.2.1 Interferometric PM Signals ............................................................50
  3.3 Experiment ...........................................................................................52
    3.3.1 The Step-Scan FTPM Arrangement ..............................................53
3.3.2 InGaAs/InGaP Quantum Well ............................................. 54
3.3.2.1 Experimental Arrangement ........................................... 54
3.3.2.2 Results and Discussion .............................................. 55
3.3.3 Bulk GaSb ................................................................ 56
3.3.3.1 Experimental Arrangement ........................................... 56
3.3.3.2 Results and Discussion .............................................. 57
3.3.4 Bulk InAs ................................................................. 58
3.3.4.1 Experimental Arrangement ........................................... 58
3.3.4.2 Results and Discussion .............................................. 59
3.3.5 Bulk InSb ................................................................ 60
3.3.5.1 Experimental Arrangement ........................................... 60
3.3.5.2 Results and Discussion .............................................. 61
3.3.6 Conclusions and Future Work ........................................... 68

3.4 References ................................................................ 70

4 Bandgap Reduction in Dilute Nitride Materials Based on Narrow Gap
Semiconductors ................................................................ 73
4.1 Introduction ................................................................ 73
4.2 Theory ......................................................................... 74
4.2.1 The Band Anti-Crossing Model ................................. 74
4.2.2 GaSb$_{1-x}$N$_x$ ............................................................. 81
4.2.3 InAs$_{1-x}$N$_x$ .............................................................. 82
4.3 Experiment I: GaSb$_{1-x}$N$_x$ ............................................. 84
4.3.1 Samples .................................................................. 84
4.3.2 Experimental Procedure ............................................. 85
4.3.3 Results and Discussion .............................................. 86
4.3.3.1 Bulk GaSb ............................................................. 87
4.3.3.1.1 Photoluminescence ............................................ 87
4.3.3.1.2 Photomodulated Reflectance .............................. 90
4.3.3.2 GaSb$_{1-x}$N$_x$ Epilayers ........................................... 93
4.3.3.2.1 Photoluminescence ............................................ 93
4.3.3.2.2 Photomodulated Reflectance .............................. 97
4.3.4 Conclusions and Future Work .................................... 97
4.4 Experiment II: InAs$_{1-x}$N$_x$ ............................................ 98
4.4.1 Samples .................................................................. 98
4.4.2 Experimental Procedure ............................................. 99
4.4.3 Results and Discussion .............................................. 101
4.4.3.1 Bulk InAs ............................................................. 101
4.4.3.1.1 Photoluminescence ............................................ 101
4.4.3.1.2 Photomodulated Reflectance .............................. 106
4.4.3.2 InAs$_{1-x}$N$_x$ Epilayers ........................................... 111
4.4.3.2.1 Photoluminescence ............................................ 111
4.4.4 Conclusions and Future Work .................................... 118
4.5 References .................................................................. 119

5 Time-Resolved Measurements of Spin Relaxation in Bulk Narrow Gap
Semiconductors ................................................................ 125
5.1 Introduction ................................................................ 125
5.1.1 The Spin-FET: A Spin-Based Semiconductor Device 125
5.2 Theory ......................................................................... 126
5.2.1 Cubic Spin-Splitting in Bulk III-V Semiconductors 126
5.2.2 Spin Relaxation Mechanisms ...................................... 133
5.2.2.1 The D'yakonov-Perel Mechanism ............................ 133
5.2.2.2 The Elliott-Yafet Mechanism .................................. 135
5.3 Experiment ................................................................. 137
5.3.1 Samples .................................................................. 137
5.3.2 Time-Resolved Optical Pump-Probe Experiments 138
5.3.3 Spin Polarisation ....................................................... 140
5.3.3.1 Faraday Rotation Method ....................................... 141
5.3.3.2 Polarisation Pump-Probe Technique ........................ 143
Abstract

A number of optical studies have been performed in this thesis. These include optical characterisation of mid-infrared semiconductor materials, high pressure investigations of mid-infrared emitting devices, and investigations of spin-dependent processes in various semiconductor alloys.

A Michelson interferometer-based photomodulation spectroscopy arrangement has been developed extending the long wavelength limit for the technique, which previously stood at a wavelength corresponding to the InAs bandgap (i.e. ~3.5μm). Photomodulated transmission measurements are demonstrated in InSb up to ~6.5μm. Dilute nitride materials have also been investigated; namely GaSbN and InAsN. While the former appears optically inert, an unambiguous redshift in the bandgap as a function of nitrogen composition has been observed in InAsN. At 300K a fit to the experimental data has yielded band anti-crossing model parameters in good agreement with those predicted by tight-binding calculations.

Spin lifetimes in bulk narrow gap semiconductors have been measured, utilising ultrafast pulses from either a free-electron laser (FELIX) or Ti:Sapphire source to optically generate and probe the decay of spin coherence in InAs, InSb and HgCdTe. Experiments on InSb include the first such measurements at elevated temperature (i.e. 300K). An increase in spin lifetime has been observed in degenerately doped InSb and InAs samples. Also, the Elliott-Yafet relaxation mechanism has been identified as the dominant relaxation path in moderately doped HgCdTe based on the spin lifetime's temperature dependence. Using the FELIX laser system, the first unambiguous observation of the spin galvanic effect in the absence of an applied magnetic field has also been made. Suppression of the accompanying circular photo-galvanic effect has been achieved by exploiting the interplay of Rashba and Dresselhaus spin-orbit coupling terms in n-type GaAs quantum wells.

Mid-infrared emitting devices based on InAs have also been investigated as a function of hydrostatic pressure. It is shown that in type-II structures Auger recombination has been identified as the dominant current path, while for the type-I device contributions from radiative recombination and Auger recombination appear more comparable.
This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that the copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author or the university (as may be appropriate).

© Martin Merrick 2006
Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Ben Murdin, for all the help, advice, and encouragement he has given throughout my time as his student. His knowledge and experience have been invaluable in the work presented in this thesis.

Dr. Jeff Hosea has also provided a great deal of assistance, particularly with regards to photomodulation experiments. Performing the role of co-supervisor in fact, if not on paper, the help encouragement provided by him with respect to the FTIR modulation spectroscopy experiments presented was much appreciated.

Of my fellow students at Surrey, there are numerous people deserving of recognition. Stuart Cripps provided both many laughs and valuable scientific contributions during the mid-infrared photomodulation experiments we worked on together. It is now his job to continue development of the FTIR photomodulation spectroscopy arrangement, which I wish him luck with. Also, Konstantin Litvinenko was integral to the spin lifetime measurements performed at Surrey, with his command of a Ti:Sapphire laser without peer! Dr. Stelios Choulis was extremely helpful during the early days of my PhD, and was an important part of the high-pressure work presented here. The department’s resident Mr. Fix-It Gary Strudwick deserves special thanks for his cheerful disposition in the face of an onslaught of technical glitches for him to repair, while the additional technical support of Dr. Nasir Ahmad was also much appreciated. Dr. Damien Clarke and Dr. Andrew Hollingworth, my predecessors as Ben’s student, were a great deal of help early on also. And I would like to thank all those I do not have space to name individually for their friendship.

Elsewhere, I would like to thank Dr. Pawel Murzyn from Heriot-Watt for his assistance with spin lifetime measurements. During experiments carried out at the FELIX Free Electron Laser Facility in the Netherlands, Dr. Jonathan Phillips provided sound technical guidance, while the other resident staff members at the facility provided vital assistance also. Dr. Louise Buckle and Dr. Tim Ashley from Qinetiq were responsible for many of the samples studied in this work, and deserve recognition for their expertise in that area. Prof. Tony Krier (Lancaster University)
and Prof. Sergey Ganichev (University of Regensburg) are also thanked for their help with work on InAs-based devices and the spin galvanic effect respectively. Correspondence with Dr. Tim Veal of the University of Warwick, and Prof. E. O’Reilly and Dr. A Lindsay of the Tyndall National Institute in the Republic of Ireland, also provided many illuminating discussions regarding dilute nitride materials. Grateful acknowledgement is also made of the funding and financial support provided by the EPSRC (Engineering and Physical Sciences Research Council).

My friends beyond Surrey did a great job of keeping me sane for the duration of this work. They were always willing to listen to whatever bizarre area of physics I decided to lecture them about, and didn’t hit me once as a result! Most of all, I thank them for their non-physics related contributions, as their friendship and good humour lifted me from PhD-related doldrums on countless occasions.

Last, but by no means least, special thanks must go to my family. The tireless support and unending patience of my parents, Barbara and Dennis, and brother, Kevin, is a primary contributor to my being able to write this thesis. Without their help, financial and personal, this simply would not have been possible, and for the opportunities they have given me I will be eternally grateful. Thank you.
Publications


1 Introduction

1.1 Optical Characterisation in the Infrared

1.1.1 Background

For many years research into the optical properties of semiconductor materials has tended to focus on particular material alloys whose fundamental energy gaps (or bandgaps) lie within a relatively narrow energy range, corresponding to wavelengths from the ultraviolet (UV) at ~0.1\(\mu\)m through to the near infrared (NIR) at ~3\(\mu\)m (and inclusive of the visible region). A driving force for this concentration of resources are the numerous commercial applications for devices operating within that range. For example, the single largest industry application of semiconductor lasers is optical media, with every PC CD-ROM, CD drive, and DVD player containing at least one light emitting diode (LED) or laser device operating at visible wavelengths. Next largest is the telecommunications industry, for which laser devices emitting within the peak transmission windows of conventional optical fibres are required (e.g. at ~1.3-1.5\(\mu\)m for glass fibres, and 0.68-1.3\(\mu\)m for plastic fibres). As customer demand for greater bandwidth and connection speeds at lower cost continues to grow unabated, so too must the development of ever faster, more powerful, efficient, and (preferably) tuneable laser devices.

Recently, however, considerable interest has developed in semiconductor materials and devices capable of emitting and absorbing photons at mid-infrared (MIR) wavelengths, between 3-20\(\mu\)m; the domain of narrow gap semiconductors (NGSs).
This growth is driven by the realisation that efficient emitters in that regime have the potential to provide new solutions to old problems, both military and commercial. For example, the former’s most publicised application is as a counter-measure to infrared (IR) sensitive devices. Specifically, an enemy soldier using IR (or night-vision) goggles could be temporarily blinded by a hypothetical high power MIR laser. Similarly, a heat-seeking missile tracking an aircraft’s engines could be effectively neutralised by a MIR laser capable of providing a far brighter heat source than the engines themselves. The most pressing commercial application lies in the relevance of MIR wavelengths to molecular bonding. In principle, a device emitting at a wavelength corresponding to the unique vibrational or rotational modes of a particular molecular species could be used as an atmospheric gas sensor.

Figure 1.1 shows the atmospheric transmittance over a broad range of the electromagnetic (EM) spectrum (UV to microwave), with emphasis on the MIR region. Currently coverage of MIR wavelengths by semiconductor alloys (as defined by their bandgap, $E_g$) is patchy at best. In terms of bulk materials, the best options for detectors remain GaSb ($E_g \sim 1.7\mu m$ at 300K, so technically NIR), InAs ($E_g \sim 3.5\mu m$ at 300K), and InSb ($E_g \sim 7.1\mu m$ at 300K). In order to assist crystal growers in their efforts to produce more exotic alloys (or structures) capable of providing emitters within the spectral gaps (i.e. 1.7-2.8\mu m, 3.5-5\mu m, >7.1\mu m), optical characterisation of grown samples is necessary. That is, key material parameters such as the bandgap, effective masses, and band-offsets must be determined in order to identify any deficiencies and suggest improvements. Unfortunately, the narrow bandgaps of MIR emitting materials also tend to make for intrinsically poor optical properties relative to those of wide gap semiconductors (WGSs), due to the presence of strong non-radiative recombination paths. These shortcomings can complicate the interpretation of optical measurements (if any can be made at all), and only become more acute in untried and untested experimental alloys such as the NGS-based dilute nitrides studied in chapter 4.

---

1 Electrical characterisation also provides valuable information on carrier concentrations, mobilities, recombination paths etc. However, the work presented in this thesis focuses primarily on optical techniques.
Figure 1.1 Atmospheric transmittance across a broad range of the electromagnetic spectrum along a 1nm sea level path. In the upper plot the region between UV and microwave wavelengths is shown, with the contributions of both absorption and scattering losses illustrated. Absorption occurs in discrete bands, corresponding to the vibrational and rotational modes of molecules such as water (H₂O) and carbon dioxide (CO₂), whereas scattering (also referred to as Rayleigh scattering) is a smoothly varying function of wavelength. The lower plot expands the infrared region of the spectrum, with near-infrared and mid-infrared regimes indicated.[15,16]

In the NIR/MIR transitional region (i.e. 1.5-3.5μm), care must be taken to choose the correct instrument for a particular spectroscopic measurement. Grating-based spectrometers and Fourier transform infrared (FTIR) interferometers are both commonly used in the neighbourhood of this cross-over, although the two instrument types do have certain advantages and disadvantages which can predispose their application to certain techniques.
1.1.2 The Czerny-Turner Grating Spectrometer

1.1.2.1 Principle of Operation

It is generally accepted that in the visible/NIR regime grating spectrometers are the instrument of choice. A commonly encountered design is the Czerny-Turner grating spectrometer, illustrated in Figure 1.2.[1]

![Diagram of Czerny-Turner Grating Spectrometer]

Figure 1.2 The key components of a Czerny-Turner grating spectrometer. Light from the broadband source is collimated before being directed onto a rotatable diffraction grating. The reflected light is then re-focused onto slit $S_2$. By rotating the grating the diffraction-generated spectrum may be scanned across $S_2$, thereby allowing a range of monochromatic wavelengths to be produced. The widths of $S_1$ and $S_2$ can be altered to change the resolution and light throughput.

Light from a broadband source is focussed onto the adjustable slit, $S_1$. This light is then collected by a collimating mirror, which directs the light onto a rotatable grating. The beam is thus split into a continuous spectrum of reflected beams, which are collected and focussed by a second mirror onto the plane of slit $S_2$. The output wavelength may be chosen by rotating the grating, thereby scanning the projected spectrum across $S_2$ until the desired wavelength falls on its centre.

Though the number of grooves illuminated by the source imposes a fundamental limit on the resolution of the spectrometer, a further user control is provided in the widths of $S_1$ and $S_2$. By reducing the size of $S_1$, the quality of the focus may be improved, creating a more point-like source at the entrance of the spectrometer. Consequently, the first mirror will be better able to collimate the beam. This is advantageous because
Introduction

Poor collimation of the beam incident on the grating results in a smearing effect on the generated spectrum, effectively producing many overlapping spectra, each offset by some small amount. However, this improved collimation comes at the expense of source intensity, reducing the signal-to-noise ratio (SNR) in an experiment. Similarly, $S_2$ places a restriction on the exit of the spectrometer, allowing only a narrow region of the projected spectrum to pass. The smaller $S_2$ is made, the narrower the bandwidth of the light exiting the spectrometer, and the greater the spectral resolution achievable.

Note that the particular arrangement shown in Figure 1.2 provides a means of generating monochromatic light over a wide range of wavelengths (i.e. the bandwidth of the source). However, by replacing the light box with some other polychromatic source of interest, for example a semiconductor sample emitting photoluminescence, an emission spectrum may be measured.

1.1.2.2 Advantages and Disadvantages

The simplicity of the Czerny-Turner spectrometer on the whole makes it a very easy instrument to work with. However, it is not without drawbacks. First to note is the 1:1 scaling of a grating's size with wavelength for a given resolving power. Hence, as wavelength is increased, gratings with high resolving powers rapidly become cumbersome and expensive. A second deficiency is the aforementioned trade-off between resolution and SNR associated with the widths of $S_1$ and $S_2$. Finally, the appearance of higher-order diffraction patterns in an output spectrum can present a significant problem, particularly at longer wavelengths. If the detector is sensitive to the higher orders present in the spectrometer throughput, the signal-to-noise ratio (SNR) can be negatively impacted. Furthermore, false images of features occurring at the higher order wavelengths can appear in the spectrum at the wavelength of interest, possibly obscuring any real features which may be there. For this reason, order sorting filters are usually used to provide what is known as free spectral range; a wavelength range in which there is no overlap of multiple orders. Although a trivial piece of equipment for visible/NIR wavelengths, as with most optics the range of MIR options is severely limited. Due to a lack of demand and high cost of machining suitable materials, a typical set may come with only a small selection of possible pass-bands. Therefore, it is not uncommon to find a situation where a feature of interest straddles
the boundary between two filter options, with neither one being an appropriate choice. Also, if a feature shifts continuously as a function of some experimental variable (e.g. temperature, pressure, etc.) a filter may need to be swapped mid-experiment.

1.1.2.3 Systems Used in this Work

Two different spectrometers, both of Czerny-Turner design, were used in experiments presented in this thesis. The highest specification model was a Jobin Yvon Triax Series 320. This spectrometer had three selectable gratings, each 2.5" square, the details of which are provided in Table 1.1.

<table>
<thead>
<tr>
<th>Grating Blaze (µm)</th>
<th>Groove Period, g (µm)</th>
<th>Operational Range (µm)</th>
<th>g/mm</th>
<th>Dispersion (nm/mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>2</td>
<td>1-3</td>
<td>600</td>
<td>5.28</td>
</tr>
<tr>
<td>5.0</td>
<td>8</td>
<td>3-10</td>
<td>150</td>
<td>2.12</td>
</tr>
<tr>
<td>15.0</td>
<td>20</td>
<td>10-30</td>
<td>60</td>
<td>52.8</td>
</tr>
</tbody>
</table>

Table 1.1 Details of the higher specification Jobin Yvon Triax Series 320 Spectrometer utilised in this work.

The lower specification model, also manufactured by Jobin Yvon, had a single 1.5" square grating, the details of which are listed in Table 1.2. This was used largely for convenience, when the higher specification version was unavailable.

<table>
<thead>
<tr>
<th>Grating Blaze (µm)</th>
<th>Groove Period, g (µm)</th>
<th>Operational Range (µm)</th>
<th>g/mm</th>
<th>Dispersion (nm/mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>4</td>
<td>0.8-3200</td>
<td>300</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 1.2 Details of the lower specification Jobin Yvon spectrometer used in this work.
1.1.3 The Michelson Design FTIR Interferometer

1.1.3.1 Principle of Operation

As one makes the transition from NIR to MIR wavelengths, a grating system becomes less practical for the reasons detailed in the preceding section. Conversely, FTIR spectrometers are less practical at visible/NIR wavelengths, but in the MIR are widely accepted as being the best option. This is certainly true for relatively simple applications such as photoluminescence (PL) or transmission spectroscopy. The most common FTIR instrument found in research laboratories is probably the Michelson interferometer, which is illustrated in Figure 1.3.[2]

Figure 1.3 The key components of a Michelson FTIR interferometer. Light from a broadband source is collimated before being incident on the beamsplitter. The length of one of the resulting beam paths is adjustable by means of moving mirror $M_1$. The retro-reflected beams (from $M_1$ and $M_2$) are subsequently recombined at the beamsplitter, with the resulting interference (or fringe) pattern as function of the beam’s path difference related to the source spectrum by a Fourier transform (see appendix, section 9.2.1).
Light from a broadband source is focused onto an aperture, $A_1$, at the entrance to the instrument. This light is then collimated, usually by a spherical or parabolic mirror. By reducing the size of $A_1$ the quality of the source focus can be improved until it is virtually a point source, thus assisting collimation at the cost of source intensity. Note that in this case the restriction is a circular aperture, rather than the slit found in Czerny-Turner spectrometers (i.e. $S_1$). Therefore, unlike in the grating case where the focus of the output beam has dimensions related to the slits (i.e. a rectangle), from an FTIR a circular focus can be produced. This makes for a more efficient use of the available light, as detector elements (for example) are generally not rectangular.

Following collimation, the beam is usually incident on the key component of a Michelson Interferometer, the beamsplitter. This usually consists of two plates (or substrates), made from a material transparent at the wavelengths of interest. The lower side of the upper plate is given a semi-silvered coating (as depicted in Figure 1.3) such that it both partially reflects and transmits in a 50:50 ratio (ideally). The incident beam is thus split into two components which travel along different paths (or arms) of lengths $L_1$ and $L_2$ respectively. Each arm is terminated by a mirror, $M_1$ and $M_2$, which reflect the beams back along a reciprocal path toward the beamsplitter. While $M_2$ is fixed relative to the beamsplitter, $M_1$ is able to translate back and forth along the beam path, thereby making $L_2$ variable. Upon their second incidence with the semi-silvered surface, the beams once again undergo partial reflection and transmission, the net effect being the 'recombination' of the two beams along a new fourth path (towards the exit of the instrument). Note that during this second partial reflection/transmission some losses are expected due to a component of each beam also travelling back along the original path (i.e. towards the source).

Because the recombined beams overlap spatially they interfere with one another, generating a fringe pattern dependent on the relative phase of each wavelength component. A phase difference may be expressed in terms of the difference in path length in each arm (i.e. in cm), or in terms of a time delay added to the round trip time of one of the beams (i.e. in s). If the paths $L_1$ and $L_2$ are precisely equal, the phase of every wavelength component in the recombined beams will be the same in each arm.

---

2 This double plate arrangement is necessary to phase match the recombined beams by ensuring they both pass through the same thickness of beamsplitter substrate material during their round trip.
(i.e. their relative phase difference is zero, as is the path difference and time delay). The resulting interference pattern is, therefore, entirely constructive in nature irrespective of wavelength. This particular mirror position is an important reference point for interferometric measurements, and is often referred to as the 'point of stationary phase' or 'zero path difference' (ZPD).

If $M_2$ is initially positioned at ZPD and moved at a fixed velocity in either direction, the fringe pattern measured at the output encodes the phase of every wavelength component as a function of path difference (or retardation). This is equivalent to the Fourier transform of a grating spectrometer's output, and is referred to as an interferogram. For a more detailed discussion of interferograms, the reader is directed to the appendix of this thesis (section 9.2).

1.1.3.2 Advantages and Disadvantages

FTIR interferometers have several advantages over grating systems. Firstly, the spectral resolution of an acquired interferogram is related to the maximum retardation. Therefore, to achieve a higher resolution one need only scan the moving mirror farther away from ZPD. At MIR wavelengths, a reasonable retardation is of the order of millimetres, so extensions to a scan's range of that order are quite trivial. Certainly it is a more favourable proposition than doubling the dimensions of a typical grating. Furthermore, unlike closing slit $S_2$ on a grating spectrometer as a means of increasing resolution, scanning farther from ZPD does not detrimentally affect the SNR.

Spectra acquired by FTIR means are also immune to the effects of higher order wavelength components. So long as the sample rate is sufficiently high, all spectral components should be uniquely resolvable. However, therein also lies an FTIR spectrometers biggest weakness, for the ability to sample on an appropriately small length scale rapidly becomes impractical for wavelengths shorter than $\sim 1\mu m$. For this reason, they are not generally used for measurements in the visible or short wavelength end of the NIR; too low a sampling rate results in aliasing of spectral components, thereby rendering spectra indecipherable (see appendix, section 9.2.2).
1.1.3.3 Systems Used in this Work

Two Michelson interferometers have been used in experiments presented in this thesis. One of these was a commercial specification Bomem MB-100 fast scanning instrument, fitted for MIR operation. This interferometer is referred to as fast scanning because the moving mirror scans back and forth continuously with a frequency (in this model) of order ~1Hz (corresponding to a mirror velocity of ~1.9mms⁻¹). Hence, it can acquire around one scan per second in normal operation. Its potassium bromide (KBr) beamsplitter’s transmission range effectively defines the instruments upper and lower wavelength limits. The quoted limits of that component are ~2μm to ~22μm, although PL signals have been obtained from a semiconductor sample emitting at 1.55μm using the instrument, indicating those limits to be soft. The minimum wavelength the instrument can resolve is 1.266μm, precisely twice the helium-neon (HeNe) laser line wavelength. This is a consequence of the HeNe calibration laser used to count fringes and synchronise data sampling, and corresponds to a sample rate of one point per HeNe fringe.

The second system was a refitted Bomem DA3 spectrometer, adapted to operate in a step-scan mode. Its application in MIR photomodulation (PM) spectroscopy measurements is the subject of extended discussion in a future chapter. It will therefore not be discussed further at this point.
1.2 Optically Probing Electron Spin Dynamics

1.2.1 Background

Traditional electronic devices store information in the electronic charge. However, during recent years significant interest has been generated in the field of spin-based electronics (or spintronics), where information is stored in the electron spin itself. The main advantages of such technology include non-volatility, increased device speed, decreased power consumption and greater integration densities (i.e. further miniaturisation).[3] The operation of a typical device is based on the alignment of carrier spins (electrons or holes) with respect to a reference direction, defined, for example, by an applied magnetic field or the magnetization orientation of a ferromagnetic film. A spin polarisation is then detected as fluctuations in another dependent quantity (usually the electric current).

Although the (multidisciplinary) field of spintronics is still in its infancy the first generation of spin-based devices have already entered the commercial market, the basic building blocks of which are the spin-valve and magnetic tunnel junction (MTJ). The discovery of the principle underlying the operation of these structures is widely credited as heralding the birth of spintronics. Referred to as the giant magnetoresistive effect (GMR), the researchers involved were awarded the 1998 Nobel Prize for Physics, 10 years after their pioneering work was published.[4] The GMR is observed in structures of alternating ferromagnetic and non-magnetic conducting layers, where the resistance of the non-magnetic layer is found to be lowest when the magnetization of the ferromagnetic layers are aligned, and highest when anti-aligned. This observation is found to hold true for currents both parallel and perpendicular to the layer interfaces. It is the origin of GMR that gives these structures their spintronic credentials, with an interaction between the spin of electrons in the non-magnetic layer and the magnetisation of the ferromagnets found to be responsible. Commercial applications include magnetic field sensors, galvanic (or current) sensors, hard drive read heads and magnetoresistive random access memory (MRAM). The latter in particular shows great potential, offering non-volatile solid state RAM with storage capacities currently in the Megabyte (MB) regime and growing.[3] It is noted, however, that none of this current generation of spintronic devices make use of
Looking to the future, it is hoped the next generation of spintronic devices will successfully incorporate semiconductors to produce a variety of devices, including the spin-FET (field-effect transistor), the spin-LED, optical switches operating at terahertz frequencies, modulators, and encoders and decoders of quantum bits for communication and computation. However, before most of these can be practically realised, there are many issues needing to be addressed by experiment and theory.[5] These include the optimisation of spin lifetimes, coherent transport of spin-polarised carriers across the length scales relevant for applications, and the means to manipulate spins on sufficiently short timescales. To that end, in order to better understand the transient nature of spin-related processes a probe operating on the same short timescales is needed, demanding, therefore, a temporal resolution of order \( \approx \text{1ps} \). One class of devices capable of fulfilling this criteria is that of the ultrafast laser. Two such systems have been used to observe the dynamic behaviour of polarised electron spins in this thesis; a free electron laser and a Ti:sapphire laser.

### 1.2.2 The FELIX Free Electron Laser System

First invented, and subsequently demonstrated, in the 1970's by John Madey and colleagues at Stanford University, the free electron laser (FEL) was a focus of a great deal of research effort in the 1980's as part of Ronald Reagan's "Star Wars" project.[7,8,9] The aim then was to use the intense infrared radiation they can generate to heat ballistic missiles to the point of destruction mid-flight. In the post-Cold War era, the technology which has emerged as a result has found a plethora of non-military applications, ranging from medical research to fundamental studies of condensed matter. FELs have indeed proven themselves as versatile scientific instruments, capable of generating pulses of intense, coherent, and tuneable radiation on very short timescales, with more than thirty such facilities now established around the world.[10]

The Dutch FEL system FELIX (Free Electron Laser for Infrared eXperiments) has been employed for experiments in this thesis.[11] Situated near Utrecht in The Netherlands, this was one of the first such facilities located in Europe, with usership...
of the system now divided among several groups from a number of different countries.

1.2.2.1 Principle of Operation

As its name suggests, an FEL uses electrons which are not bound to atomic nuclei in its normal operation. A heated cathode, or electron gun, is used as a source of such electrons, which are then accelerated to relativistic velocities by a LINAC (LINear ACcelerator) in preparation for insertion into the lasing cavity. This cavity is surrounded by a series of magnets arranged in an alternating N-S, S-N configuration, as illustrated in Figure 1.4. Due to the Lorentz force, electrons passing through the resulting magnetic field trace an undulating path, exhibiting a transverse "wiggling" behaviour (the undulator magnets are often referred to collectively as the wiggler). This transverse component of the electron's motion induces the spontaneous emission of dipole radiation, the frequency of which is dependent on the longitudinal spacing between magnets and the velocity of the electrons in question.

Because the electrons move at close to the speed of light, the effects of special relativity must be taken into account. In particular, time dilation results in a Doppler shift of the emitted wavelength (as viewed in the lab frame), while the effect of Lorentz contraction reduces the apparent spacing between undulator magnets (as...
viewed from the electrons rest frame). Both phenomena result in a blue-shift of the emission wavelength in the lab frame, the shift being characterised by the Lorentz factor, $\gamma$. The emitted photon wavelength may therefore be expressed as follows:

$$\lambda_0 = \frac{\lambda_u (1 + K^2)}{2\gamma^2}$$  \hspace{1cm} (1.1)$$

where $\lambda_0$ is the wavelength of the spontaneous emission, $\lambda_u$ is the period of the undulator magnets, and $K$ is proportional to the magnetic field strength. For FELIX the attainable photon wavelengths lie in the IR range, between 4.5\,\mu m-250\,\mu m.[12]

The endpoints of the lasing cavity are defined by two mirrors, one of which is partially transmitting to allow extraction of the laser light. The spacing between these mirrors is referred to as the cavity length. As an electron's spontaneous emission reflects back and forth within the cavity, if the phase of the resulting optical field matches that of an injected electron's transverse motion stimulated emission occurs, and a condition for optical gain is realised. However, in emitting spontaneous emission an electron loses energy. The resulting retardation shifts an electron out of phase with the optical field, in which case it may regain that lost energy and shift back into phase by absorbing a photon from the optical field. This constitutes a loss mechanism. Saturation of the gain is said to occur when the rate at energy is dissipated by various loss mechanisms (including absorption) equals the rate of stimulated emission. For FELIX, this occurs when the stimulated emission rate is around 10 to 100 million times that of spontaneous emission alone.

**1.2.2.2 Capabilities**

The pulse structure of the stimulated emission is essentially the same as that of the electrons injected into the LINAC. In order to facilitate the dissipation of heat which builds up in the walls of the LINAC, electrons are injected only in short bursts, leading to a discrete pulse-like structure in the laser emission. Referred to as “macropulses”, these are in turn sub-divided (by design) into still shorter pulses called “micropulses”. The FELIX pulse structure is illustrated in Figure 1.5. The typical
duration of a micropulse is fixed at around 1ps, though it is possible to change their repetition rate within a macropulse, as well as altering the temporal spacing of macropulses themselves. A summary of typical pulse parameters is provided in Table 1.3. It may be noted that by altering the cavity length it is possible to detune the phase of the optical field with respect to an electron’s undulations, thereby changing the pulse profile. This is referred to as desynchronisation, with a large phase difference producing broad pulses with low peak powers, and vice versa.[13]

![Figure 1.5](image)

**Figure 1.5** The pulse structure of laser radiation emitted by FELIX. Micropulses of ~1ps duration form discrete trains of macropulses ~10μs long. Typically macropulses are separated temporally by a few hundred milliseconds, with micropulses within the macropulses arriving ~20ns apart.

Using free electrons in an evacuated chamber as its lasing medium affords an FEL numerous advantages over conventional laser systems. Firstly, the gain medium has approximately the same refractive index as air. The emitted radiation can, therefore, be perfectly linearly polarised, with ideal diffraction limited dispersion. High peak powers may also be obtained, as the normal limitations imposed by heating of the gain medium in high power lasers do not apply. Any unused electron energy is instead carried away to an external electron dump.

FELIX in particular also has a high degree of tuneability, able to change its emitted wavelength by as much as two orders of magnitude in minutes. Demonstration of such rapid tuning capabilities in an FEL was in fact a FELIX design objective.[14]
Introduction

For this work's application, perhaps the single most important advantage offered by FELIX is the ability to emit picosecond pulses with sub-picosecond pulse jitter. This follows from the fact that the radiation pulses are correlated with the electron pulses, with the latter under relatively precise user control by virtue of the well understood LINAC component of FEL technology.

<table>
<thead>
<tr>
<th>Wavelength Range</th>
<th>4.5 to 250µm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Micropulse Energy</td>
<td>50µJ</td>
</tr>
<tr>
<td>Micropulse Repetition Rate</td>
<td>1GHz, 50MHz or 25MHz</td>
</tr>
<tr>
<td>Micropulse Repetition Time</td>
<td>1ns, 20ns or 40ns</td>
</tr>
<tr>
<td>Macropulse Repetition Rate</td>
<td>≤10Hz</td>
</tr>
</tbody>
</table>

Table 1.3 A selection of possible FELIX pulse parameters.

1.2.3 The Coherent Ti:Sapphire Laser System

1.2.3.1 Principle of Operation

In order to generate ultra short-pulses, it is necessary for that pulse to be spectrally broad. This is a consequence of Fourier analysis, a subject discussed in more detail in the appendix in the context of interferograms (section 9.2). The gain medium of a laser designed to generate ultra-short pulses must, therefore, have a very large gain bandwidth. That is, it must be able to uniformly amplify all the component wavelengths of a broad stimulated emission spectrum. Titanium-doped sapphire (or Ti:sapphire) crystals are found to possess the largest bandwidth of all commonly used lasing media, as is shown in Table 1.4. Titanium-doped aluminium oxide (Ti:Al₂O₃) (to give its full name) is, therefore, a favoured option for solid state laser systems designed to generate ultrafast pulses (with length of the order of femtoseconds).
Lasing Medium & Gain Bandwidth (nm) 
---
Argon-ion & 0.007 
Ruby & 0.2 
Nd:YAG & 10 
Dye & 100 
Ti:sapphire & 400 

Table 1.4 Gain bandwidths of some common lasing media.[17]

In the Ti:sapphire crystal structure, Ti$^{3+}$ ions substitute for Al$^{3+}$ ions at concentrations up to around 0.1% by weight.[16] The absorption band of the resulting material lies in the blue-green region of the visible spectrum, and is very broad, lying between ~400nm and ~650nm. Pumping of the crystal is usually achieved using a powerful blue-green solid state laser (e.g. argon-ion, Nd:YVO$_4$), typically capable of generating several watts of continuous (or cw) emission. A Ti:sapphire crystal’s emission spectrum is also similarly broad, though shifted towards longer wavelength, covering a broad range between ~600nm and 1050nm. The emission and absorption bands are shown in Figure 1.6. The broadness of the emission peak thus allows either tuneable emission over a large range, or more importantly for the work presented in this thesis, enables the generation of ultrashort pulses.

![Absorption and emission bands of a Ti:sapphire crystal. Both lie within the visible/NIR ranges. The width of these bands is the key to a Ti:sapphire crystal’s ability to generate ultrashort pulses.][18]
1.2.3.2 Capabilities

The layout of the Coherent Ti:sapphire laser system at the University of Surrey is shown schematically in Figure 1.7. The system as whole is made up of several discrete segments, distributed over the surface of an optical bench. The first step in ultrafast pulse generation is optical pumping of the oscillator by one of the solid state Nd:YVO$_4$ lasers. At this stage the Ti:sapphire crystal (inside the oscillator) is used to generate a ‘seed’ pulse. Already these pulses are typically <100fs long, but are of limited power and repetition rate. The seed pulse is fed into the stretcher/compressor module, which on this occasion stretches the pulse, resulting in a longer pulse with lower peak power. This stage is necessary to prevent damaging certain optical components in the amplifier through heating effects. The subsequent output of the amplifier is fed back into the stretcher/compressor once more, where the reverse operation of compression is performed. The peak power after this stage is typically over 1000 times larger than the output from the oscillator (increased from ~65kW to ~80MW). This pulse then passes to the Optical Parametric Amplifier (or OPA) where it is split into two beams called the ‘signal’ and the ‘idler’. One can use one or both of these beams for experiments, each having component wavelengths 1.1-1.6µm (signal) and 1.6-2.4µm (idler) respectively. Alternatively, this output can be forwarded to a final stage called the Difference Frequency Generator (or DFG). As its name suggests, at this stage the signal and idler pulses are mixed to produce a beam with a frequency equal to their difference. This allows generation of long wavelength ultrashort pulses in the range 2.5-10µm (the wavelengths of interest in this thesis). A summary of some of the key beam parameters at each stage can be found as annotations in Figure 1.7.
Introduction

Figure 1.7 Schematic providing an overview of the Coherent Ti:sapphire laser system at the University of Surrey. Use of the DFG to generate mid-infrared femtosecond pulses was the laser's ultimate application in this thesis. Various pulse/beam parameters are shown for a number of points along the pulse's route.

1.3 References


Introduction


2 Optical Characterisation Techniques

2.1 Introduction

Knowledge of the electronic bandstructure of semiconductor alloys as a function of variables such as composition and heterostructure combination is of vital importance for material and device engineering. Bandstructure is usually parameterised semi-empirically, and various optical techniques are available for measuring those parameters. In this thesis two optical characterisation techniques have been employed; photoluminescence and photomodulation spectroscopy.

Photoluminescence, as observed in bulk semiconductors, is the radiative emission of photons due to the recombination of electron-hole pairs generated via optical excitation. Although the processes are intimately related, the widths of the features associated with PL are considerably narrower than those observed in absorption, making PL a more reliable tool for accurately determining various transition energies (including $E_g$). In addition, PL is arguably among the simplest optical experiments to perform and analyse. However, in bulk materials only transitions associated with the fundamental bandgap and deeper-lying impurities are usually observed above 4K.[1]

All modulation spectroscopy techniques, of which there are a considerable number of variations, share a common fundamental principle; ultimately it is the (periodic) change induced in a solid’s dielectric constant due to an applied perturbation that is measured via some dependent variable. This dependent variable may fall into one of two groups; an ‘internal’ modulation or an ‘external’ modulation. The former includes
the modulation of a parameter within the measurement system, such as the wavelength or angle of incidence of a probe beam reflected from the sample surface. External modulation refers to perturbations applied directly to the sample itself. These may include changes in temperature, stress, or internal electric field (among others).[2]

In the case of photomodulation spectroscopy it is the sample’s internal electric field which is modulated. In this work, the photoreflectance (PR) and phototransmittance (PT) variants of the PM technique have been employed. These have proven valuable tools for analysing the bandstructure of semiconductor materials and heterostructures in the wavelength regime lying between the UV and NIR (i.e. ~0.15-2μm). Of particular note is the pivotal role PR has played in developing the theory of dilute nitride bandstructure (see section 4.2).[3] The technique’s success at these wavelengths (in particular with respect to GaAs and related material systems) stems from three main advantages; it is non-destructive (i.e. contactless), sensitive to transitions involving most of the key regions in a materials bandstructure (i.e. the critical points, including excited states), and capable of achieving room temperature resolutions rivalling those of PL at 4K.[2,4,5]

What follows is intended to provide an introduction to the theory underpinning each technique. Some key advantages and disadvantages of both will be noted, and a very general overview of their practical implementation will also be presented.

### 2.2 Photoluminescence

#### 2.2.1 Theoretical Background

The relationship between emission and absorption was first identified by Einstein. In his work on the subject he introduced what have since become known as the Einstein coefficients, $A_{mn}$ and $B_{mn}$, representing the transition rate per unit electromagnetic energy density for spontaneous emission and absorption (or equivalently stimulated emission) between non-degenerate levels $m$ and $n$. These coefficients are linked by the following expression:[6]
and the constant of proportionality is the energy density of electromagnetic modes of frequency between \( \nu \) and \( \nu + d\nu \) (or photon energy density, \( \rho_\nu \)) inside the medium, \( n_r \) the medium’s refractive index, \( h \) Planck’s constant, and \( c \) the speed of light in a vacuum. For a semiconductor in thermal equilibrium, the principle of detailed balance requires that the emission and absorption rates between valence and conduction band states separated by (photon) energy \( h\nu \) must be equal. This leads to the van Roosbroeck-Shockley relation:[6]

\[
P_{ve}(\nu)\rho(\nu) = R_{cv}(\nu) \tag{2.2}
\]

with \( P_{ve}(\nu) \) being the total absorption rate (equal to the product of the absorption coefficient and the electromagnetic wave’s group velocity, \( \alpha(\nu)c/n_r \)), \( R_{cv} \) the total emission rate, and \( \rho(\nu) \) the photon density in the frequency interval \( d\nu \). The latter is simply \( \rho_\nu/h\nu \) multiplied by the occupation function for photons. As bosonic particles, photons obey Bose-Einstein statistics, so one may write an expression relating the total rate of emission to the absorption coefficient, \( \alpha(\nu) \):[6]

\[
R_{cv} = \frac{\alpha(\nu)8\pi h^2 n_r^2}{c^2[\exp(h\nu/k_B T)-1]} \tag{2.3}
\]

where \( k_B \) is the Boltzmann constant, and \( T \) is the ambient temperature. Inputting the appropriate function for \( \alpha(\nu) \), this expression informs of the ambient emission rate as a function of photon energy for a semiconductor in thermal equilibrium (i.e. at the same temperature as the photon bath in which it resides - the Planck distribution at temperature \( T \)).

However, exciting electron-hole pairs within the sample using a laser beam drives the system away from thermal equilibrium, unbalancing the total emission and absorption rates. It is the excess photons emitted by the sample during the process of returning to thermal equilibrium that one generally detects in a PL experiment. The non-
equilibrium populations of electron-hole pairs generated by a laser excitation are often described in terms of electron and hole quasi-Fermi levels. The more intense the driving excitation, the further the quasi-Fermi levels are shifted from the material’s actual Fermi level, and the greater the total emission rate (i.e. more intense PL).

Under cw excitation, carrier-carrier and carrier-phonon interactions will cause the system to rapidly tend toward a dynamic (or quasi-) thermal equilibrium, in which the quasi-Fermi levels remain fixed (at different energies). The total emission rate from a semiconductor in this state may then be written as follows:[6]

\[ R_{cv} = A_{cv} f_c f_h \]  \hspace{1cm} (2.4)

where \( A_{cv} \) is the Einstein coefficient and is proportional to the joint density of states (i.e. the number electronic states per unit photon energy), \( f_c \) is the Fermi-Dirac distribution for electrons in the conduction band, and \( f_h \) is that for holes in the valence band. These occupation functions read as follows:[7]

\[
 f_c = \frac{1}{\exp\left(\frac{E_c - E_{fe}}{k_B T}\right) + 1}, \quad f_h = \frac{1}{\exp\left(\frac{E_{fh} - E_h}{k_B T}\right) + 1} \]  \hspace{1cm} (2.5)

where \( E_{fe} \) and \( E_{fh} \) are the quasi-Fermi energies for electrons and holes respectively, with \( E_{fe} \) zero at the conduction band edge and positive into the band, and \( E_{fh} \) zero at the valence band edge and positive into that band. Note that generally the temperature, \( T \), in equation 2.5 can be significantly larger than ambient due to the local deposition of energy into the crystal by the excitation source. Assuming parabolic bands, the joint density of states is proportional to \( (\hbar \nu - E_g)^{1/2} \) (see appendix, section 9.4), so one may write the following expression for the spectral intensity distribution of band-to-band PL emission from a semiconductor:

\[
 I_{pl}(\hbar \omega) = A (\hbar \omega - E_g)^{1/2} f_c f_h \]  \hspace{1cm} (2.6)

where constant term \( A \) includes the conduction and valence band masses, while
\( \hbar = h/2\pi \) and \( \omega = 2\pi v \). The effect of self-absorption by the sample on the measured PL spectrum has been neglected in equation 2.6. If one considers a perfectly crystalline wide bandgap intrinsic semiconductor, with parabolic bandstructure, under low excitation conditions (i.e. the quasi-Fermi levels lie close to the Fermi level, which in turn lies in the centre of the forbidden energy gap, far from the band edges, so that \( \hbar \omega >> k_B T \)), then the occupation density of the bands is small and Fermi-Dirac statistics (equation 2.5) may be approximated by a Boltzmann distribution. Both electron and hole occupations are then identically equal to \( \sim 1/\exp[\hbar \omega/(2k_B T)] \), yielding the following approximation for the PL spectral distribution:[6,8]

\[
I_{\text{PL}}(\hbar \omega) \propto (\hbar \omega - E_g)^{1/2} \exp \left( -\frac{\hbar \omega}{k_B T} \right) \tag{2.7}
\]

By differentiating this result with respect to \( \hbar \omega \), the photon energy at which the lineshape is peaked, \( \hbar \omega_{\text{max}} \), can be found:[8]

\[
\hbar \omega_{\text{max}} = E_g + \frac{k_B T}{2} \tag{2.8}
\]

Thus the actual band-to-band transition energy of PL emission from an intrinsic WGS is located approximately \( k_B T/2 \) below its peak energy. Although this idealised lineshape can be extremely successful at fitting the high energy tail of experimental features, it is unable to account for the broadening of a peaks low energy side as observed in real materials. Fitting experimental lineshapes using this distribution can, therefore, be rather awkward. However, it is common practice in the literature to evaluate the PL peak photon energy, and then simply take \( E_g \) as being \( k_B T/2 \) less.[9]

It was first noted by Urbach that the absorption coefficients of many solid state materials exhibit an anomalous low energy exponential decay.[10] This has come to be known as the Urbach tail, and in a semiconductor results in absorption at photon energies smaller than the bandgap. In an otherwise intrinsic semiconductor this tail is generally attributed to imperfections within the semiconductor's crystal lattice (e.g. impurities, dislocations, etc.), which lead to the formation of localised states within
the bandgap.[11] In crystalline semiconductors these states are usually only found close to the band edges, although in some amorphous materials they can actually completely bridge the bandgap. As localised states they have a mobility of zero, so the regions in which they are found are commonly referred to as the mobility gap (the term energy gap no longer being applicable). It is optical transitions between states in the mobility gap that gives rise to the Urbach tail, with absorption in the tail obeying approximately Urbach's rule:[12]

\[
\alpha(\hbar \omega) = K \exp \left[ \frac{\sigma}{k_B T} \left( \hbar \omega - E_g \right) \right]
\]  

(2.9)

where \( \sigma \) is a phenomenological parameter describing the slope of the Urbach tail, and \( K \) is a fitting parameter with no specific physical interpretation. Applying the appropriate continuity conditions, the energy at which equation 2.6 and equation 2.9 cross-over can be shown to be \( E_{cr} = k_B T / 2 \sigma \).[12] Thus a more practical PL lineshape than that of equation 2.6 alone may be defined as follows:

If \( \hbar \omega \leq E_{cr} \)

\[
I_{PL}(\hbar \omega) = A \sqrt{\frac{k_B T}{2 \sigma}} \exp \left[ \frac{\sigma}{k_B T} (\hbar \omega - E_{cr}) \right] f_e f_h
\]  

(2.10)

If \( \hbar \omega > E_{cr} \)

\[
I_{PL}(\hbar \omega) = A (\hbar \omega - E_g)^{\nu/2} f_e f_h
\]  

(2.11)

It may be noted that if \( T \to 0 \) (i.e. mobility gap states are neutralised) or \( \sigma \to \infty \) (i.e. there are no mobility gap states), then equation 2.10 \( \to 0 \) and \( E_{cr} \to E_g \), reducing the lineshape to the form described by equation 2.6.

So far the semiconductor material in question has been assumed to be intrinsic. By doping the material an extrinsic semiconductor may be obtained. Assuming \( n \)-type doping, the occupation density in the conduction band can eventually become large enough that the Pauli Exclusion Principle (PEP) plays a significant role in determining state occupations. When this occurs, the Boltzmann approximation of equation 2.7 breaks down, and PL emission can only be described accurately using
Fermi-Dirac statistics.[6] Changes in PL lineshape within this PEP-governed regime include broadening of the emission's high energy tail, and a shift of $\hbar\omega_{\text{max}}$ to higher energy, as illustrated in Figure 2.1. The latter is of particular importance, as incorrectly applying the Boltzmann approximation (and equation 2.8) during analysis can lead to an over-estimation of a PL feature's associated transition energy. Such occupancy related changes in lineshape are collectively referred to as 'band-filling' effects. The blue-shift in the absorption edge which accompanies band-filling is known as a Moss-Burstein Shift.[13,14]

![Figure 2.1](image_url)

**Figure 2.1** Comparison of functional forms commonly used to fit experimental PL spectra. The blue-dashed line is the simplest theoretical lineshape, and utilises the Boltzmann approximation (equation 2.7). The red-dashed line is identical but for the addition of an Urbach tail (equation 2.9), which better reproduces the low energy tail of practical features. The green-dashed line is a more accurate distribution based on Fermi-Dirac statistics (equations 2.10 and 2.11). At low carrier concentrations the Fermi-Dirac and Boltzmann approximation approaches yield equivalent results. In this example, $E_g$=0.3eV (note a narrow gap), $n$=$2\times 10^{15}$cm$^{-3}$, with $n$-type doping $-5\times 10^{16}$cm$^{-3}$ and $-1\times 10^{15}$cm$^{-3}$ photo-excited carriers. The differences in lineshape (e.g. broadening of the high energy tail, peak shift) due to band-filling effects are clearly seen, with the inset showing that $\hbar\omega_{\text{max}}$ is more than $k_BT/2$ from $E_g$ in the fermi-Dirac case. This can lead to overestimation of $E_g$ if the Boltzmann approximation is incorrectly assumed valid at high carrier densities.

If one fits equations 2.10 and 2.11 to an experimental PL lineshape in this regime, the quasi-Fermi energies obtained from the fit can (in principle) be used to extract the source semiconductor's actual Fermi level. The first step in doing so is calculation of
the carrier density in each band, which can be found from $E_{fe}$ and $E_{th}$ via the following relations:

$$n = N_c \exp \left( \frac{E_{fe}}{k_B T} \right), \quad \text{and} \quad p = N_v \exp \left( \frac{E_{th}}{k_B T} \right)$$  \hspace{1cm} (2.12)$$

where $n$ and $p$ are the carrier densities in the conduction band and valence band respectively, and $N_c$ and $N_v$ are material dependent parameters related to the density of states in each band. Assuming knowledge of the band effective masses (from literature sources) and lattice temperature (which can be obtained from a PL features high energy tail), $n$ and $p$ during the quasi-thermal equilibrium state existing at the time of the PL emission can thus be estimated. When the excitation ceases, the carrier concentrations decrease (in a 1:1 ratio) due to recombination, and the quasi-Fermi levels re-stabilise at the actual Fermi level. Rearranging equation 2.12 for the quasi-Fermi energies, this re-stabilisation may be represented by plotting $E_{fe}$ and $E_{th}$ as a function of decreasing $n$ and $p$ respectively. The actual Fermi level is then simply the energy at which $E_{fe}$ and $E_{th}$ are equal.

The larger effective mass of the valence band means far higher $p$-type doping levels are required to produce occupation densities large enough for carriers to be significantly influenced by the PEP. Therefore, Moss-Burstein shifts are more apparent in $n$-type materials. As one enters a regime of extremely heavy doping (typically $>10^{19}$ cm$^{-3}$ in WGSs) more extreme effects such as Mott transitions ensue, and PL lineshapes broaden dramatically.[6] Occurring when acceptor or donor concentrations become so high the impurity bands they occupy overlap the conduction or valence bands themselves, the onset of such a transition can manifest as a small value of $\sigma$ (e.g. $\leq 1$) in a fit to the resulting PL using equation 2.10 and 2.11.

Strictly speaking, equations 2.10 and 2.11 apply only to semiconductors whose PL emission is dominated by band-to-band recombination. In extrinsic semiconductors, at low temperature shallow donors and acceptors can become neutralised and recombination involving them may be observed also.[1] Strong mixing with the conduction band edge lends such levels an extended-like character, so the resulting
emission distributions are similar to that described by equations 2.10 and 2.11.\cite{1,6} In contrast, deeper lying impurity states are highly localised in real space, exhibiting little mixing with conduction band states. Carriers falling into an impurity's potential are thus trapped, and recombine predominantly by non-radiative mechanisms referred to as Shockley-Reed-Hall processes. Such states emit little or no PL.\cite{1}. In a high purity intrinsic semiconductor, the lowest energy states observable in low temperature PL (i.e. \(\leq 10K\)) are excitonic in nature, and are characterised by very narrow spectral distributions.\cite{1,6}

### 2.2.2 Practical Implementation

The excitation source in PL experiments is almost always a laser, making use of their high power densities and coherence. The laser beam is directed onto the sample, where it excites electron-hole pairs. Upon their recombination, the PL is emitted into a 4\(\pi\) solid angle with no memory of the orientation of the exciting photon's momentum\(^3\) (i.e. the laser beam's angle of incidence is irrelevant).\cite{1} This light is then collimated and directed into a monochromator, where it is spectrally resolved. The intensity of each spectral component is then detected, and an acquisition computer reconstructs the emission spectrum. The technique works in cw mode, though the addition of a mechanical chopper to the laser source is common. In conjunction with signal processing by a lock-in, the SNR can be significantly improved in this way.

In WGSs radiative recombination efficiencies are typically very good, so relatively low power sources (e.g. a 5mW HeNe laser) are usually adequate for the task. However, NGSs produce relatively weak PL due to the increased efficiency of non-radiative recombination mechanisms, particularly Auger recombination (see section 7.2.1) processes. Hence, in the MIR regime it is common to use considerably higher power sources (i.e. \(\geq 50mW\)) in order to achieve a greater SNR.\cite{9,15-17} However, too high an excitation power can induce excessive heating within the sample, resulting in a temperature related shift of the PL emission. To check whether this is an issue, a study of the PL emission as function of excitation power can be performed. Any such shift should thus be apparent.

---

\(^3\) This is a result of carrier thermalisation which randomises carrier momenta.
Figure 2.2 Schematic of a typical MIR PL setup. The sample is optically pumped by the excitation laser. The emitted PL (black lines) is then spectrally resolved by the monochromator, and directed onto the detector (green lines). If a mechanical chopper is used to modulate the pump laser, detector signals can be processed by a lock-in to improve the SNR. An acquisition computer re-constructs the PL spectrum (and controls the monochromator).

In anticipation of this work’s emphasis on MIR wavelengths, it should be noted that common glass-based lenses exhibit poor transmission properties in this regime, particularly beyond ~4.5μm. For this reason (90°) off-axis parabolic mirrors have been used for beam guidance, collimation, and focussing throughout, with the amount of glass in the probe beam path kept to a minimum (ideally there would be none, though circumstances do not always permit). Though generally not as versatile or simple to set up as lenses, with appropriate care good quality beam collimation and focussing can be achieved. An example PL arrangement using such mirrors is illustrated in Figure 2.2.
2.3 Photomodulation Spectroscopy

2.3.1 Theoretical Background

In photo-modulation spectroscopy a laser beam is used to alter the internal (or in-built) electric field of the sample. It does so by photo-exciting electron-hole pairs, which drift apart under the influence of the in-built field. Subsequent accumulation of carriers at the in-built field minimum or maximum (for electrons and holes respectively) serves to reduce the magnitude of the field, thereby altering the dielectric constant of the sample by an amount $\Delta \varepsilon = \Delta \varepsilon_1 + i\Delta \varepsilon_2$.[5] Focussing on the PR mode of operation (purely for the sake of argument), a bulk sample's reflectivity, $R$, is given by Fresnel's reflection equations. For normal incidence at an air-semiconductor boundary $R$ is independent of polarisation, and Fresnel's reflection equations reduce to:[18]

$$R = \left( \frac{n_r - 1}{n_r + 1} \right)^2 \quad (2.13)$$

where $n_r$ is the complex refractive index of the sample. Noting that $\varepsilon = n_r^2$ (see appendix, section 9.4), equation 2.13 may be expressed in terms of $\varepsilon_1$ and $\varepsilon_2$ as follows:

$$R = \frac{\left( \varepsilon_1^2 + \varepsilon_2^2 \right)^{1/2} - \left[ 2\varepsilon_1 + 2\left( \varepsilon_1^2 + \varepsilon_2^2 \right)^{1/2} \right]^{1/2}}{\left( \varepsilon_1^2 + \varepsilon_2^2 \right)^{1/2} + \left[ 2\varepsilon_1 + 2\left( \varepsilon_1^2 + \varepsilon_2^2 \right)^{1/2} \right]^{1/2}} + 1 \quad (2.14)$$

The change in the reflectivity, $\Delta R$, resulting from the laser induced changes $\Delta \varepsilon_1$ and $\Delta \varepsilon_2$ may be found by differentiating this expression:[19,20]

$$\Delta R = \frac{\partial R}{\partial \varepsilon_1} \Delta \varepsilon_1 + \frac{\partial R}{\partial \varepsilon_2} \Delta \varepsilon_2 \quad (2.15)$$

This change in reflectivity is fractional, so it is usual to normalise out any dependence
Optical Characterisation Techniques

of $\Delta R$ on changes in $R$ by computing $\Delta R/R$. Equation 2.15 may then be re-written:

$$\frac{\Delta R}{R} = \alpha_R \Delta \varepsilon_1 + \beta_R \Delta \varepsilon_2,$$  \hspace{1cm} (2.16)

$$\alpha_R = \frac{1}{R} \frac{\partial R}{\partial \varepsilon_1} \quad \text{and} \quad \beta_R = \frac{1}{R} \frac{\partial R}{\partial \varepsilon_2} \hspace{1cm} (2.17)$$

where $\alpha_R$ and $\beta_R$ are the Seraphin coefficients for reflection. In bulk single layer systems like those investigated in this thesis, the Seraphin coefficients are expected to be smooth, slowly varying functions of wavelength. A common approach to modulation spectra analysis expresses Seraphin coefficients in terms of unknown amplitude and phase parameters as follows:

$$(\alpha - i\beta) = Ce^{\theta} \hspace{1cm} (2.18)$$

where $C$ and $\theta$ are the amplitude and phase variables respectively. Combining equation 2.16 and equation 2.18 $\Delta R/R$ may then be rewritten:[21]

$$\frac{\Delta R}{R} = \text{Re}\left[C e^{\theta} (\Delta \varepsilon_1 + i \Delta \varepsilon_2)\right] \hspace{1cm} (2.19)$$

The Schrödinger equation for an electron in the presence of an electric field in the $z$-direction is given by:

$$\left(-\frac{\hbar^2}{2\mu} \nabla^2 - e\mathcal{E}z\right)\Psi(r) = E_{\Psi} \Psi(r) \hspace{1cm} (2.20)$$

where $\mu=(1/m_e + 1/m_h)^{-1}$ is the reduced effective mass of an electron and a hole, $e$ is the electronic charge, $\mathcal{E}$ is the magnitude of the electric field, $z$ is the displacement along the $z$-direction, $\Psi(r)$ is the electron’s wavefunction, and $E_{\Psi}$ is the electron’s total energy. It can be shown that the solutions of equation 2.20 take a form closely resembling the Airy function:[22]
\[ \Psi(\xi) = \frac{(eE)^{\frac{1}{2}}}{\hbar \Theta} \text{Ai}(\xi), \quad \text{where} \quad \xi = \frac{E_p}{\hbar \Theta} - z \left( \frac{2 \mu e E}{\hbar^2} \right)^{\frac{1}{3}} \] (2.21)

and \( \text{Ai}(\xi) \) is an Airy function. Also, \( \hbar \Theta \) is a field-dependent variable referred to as the ‘electro-optic energy’ and is given by:[22]

\[ \hbar \Theta = \left( \frac{e^2 \mathcal{E}^2 \hbar^2}{2 \mu} \right)^{\frac{1}{3}} \] (2.22)

Noting that, for direct transitions between conduction band and valence band states in a semiconductor, the square of the transition matrix element, \( |M_{cv}| \), in equation 9.20 (see appendix, section 9.4) is proportional to the square of the electron’s wavefunction, \( |\Psi(\xi=0)|^2 \), some further mathematics yields the following expression for the perturbed imaginary component of the dielectric function, \( \varepsilon_2(\hbar \omega, \mathcal{E}) \):[22]

\[ \varepsilon_2(\hbar \omega, \mathcal{E}) \approx \frac{2 \pi e^2}{m_e \hbar^2 (\hbar \omega)^2} \left( \frac{\hbar \Theta}{m_e \hbar^2 (\hbar \omega)^2} \right)^{\frac{1}{2}} - \infty \int_{\xi_0}^{\infty} \text{Ai}(\xi) d\xi \] (2.23)

where \( \xi_0 = \xi(\mathcal{E}=0) \), and \( \hbar \omega \) is the transition (or photon) energy. The integral over the Airy function in this equation replaces the joint density of states function in the zero-field situation represented by equation 9.20. In effect, the electric field removes any lattice translational invariance in the \( z \)-direction. Mixing of neighbouring Bloch states results, thereby permitting both vertical (i.e. direct) momentum conserving first-order optical transitions and, now, non-vertical (i.e. indirect) non-momentum conserving ones also. The energy surface over which this mixing occurs grows as the electric field is increased, allowing optical transitions between Bloch states with ever larger separations in momentum space to proceed. The accompanying changes in transition probabilities may be accounted for in terms of a weighted average of the zero-field density of states, which when convolved with equation 9.20, or \( \varepsilon_2(\hbar \omega, 0) \), yields equation 2.23. At photon energies where \( \varepsilon_2(\hbar \omega, 0) \) is smoothly varying this convolution has little effect, with \( \varepsilon_2(\hbar \omega, \mathcal{E}) \approx \varepsilon_2(\hbar \omega, 0) \). However, near critical points where there exist sharp discontinuities in \( \varepsilon_2(\hbar \omega, 0) \) relatively drastic changes can be
observed in $\varepsilon_2(\hbar \omega, E)$.\[23\] It is these changes one measures in PM, and for this reason it (and related modulation techniques) is often described as probing the joint density of states. Solving the integral in equation 2.23 yields the following expression for $\varepsilon_2(\hbar \omega, E)$:\[22\]

$$\Delta \varepsilon_2(\hbar \omega, E) \approx \frac{2 e^2}{m_e^2 h^2 (\hbar \Theta)^2} \left( \frac{E_g - \hbar \omega}{h \Theta} \right)^{1/2} \mathcal{F} \left( \frac{E_g - \hbar \omega}{h \Theta} \right)$$ (2.24)

where $(E_g - \hbar \omega)/h \Theta = \xi_0$, and $\mathcal{F}(\xi_0)$ is a function of $\text{Ai}(\xi_0)$. Note that $E_g$ appears in equation 2.24 assuming a transition at the $\Gamma$-point, though any critical point energy in a given bandstructure is equally valid. A Kramers-Kronig transform (equations 9.18 and 9.19 in the appendix, section 9.4) yields a similar expression for $\Delta \varepsilon_1(\hbar \omega, E)$:\[19\]

$$\Delta \varepsilon_1(\hbar \omega, E) \approx \frac{2 e^2}{m_e^2 h^2 (\hbar \Theta)^2} \left( \frac{E_g - \hbar \omega}{h \Theta} \right)^{1/2} \mathcal{G} \left( \frac{E_g - \hbar \omega}{h \Theta} \right)$$ (2.25)

where $\mathcal{G}(\xi_0)$ is related to $\text{Ai}(\xi_0)$ also, but contains additional components resembling a modified Airy function, $\text{Bi}(\xi_0)$. The functions $\mathcal{F}(\xi_0)$ and $\mathcal{G}(\xi_0)$ are illustrated in Figure 2.3.

![Figure 2.3](image_url)

**Figure 2.3** The functions $\mathcal{F}(\xi_0)$ and $\mathcal{G}(\xi_0)$, where $\xi_0 = (E_g - E)/h \Theta$. They are related to Airy functions, and contribute an oscillatory component to $\Delta \varepsilon$. 

36
Hence, in conjunction with equation 2.16, it can be seen that the general form for PM lineshapes is Airy function-like, with oscillations which decay exponentially with increasing energy for $\hbar \omega > E_g$, and an evanescent tail which decays exponentially (at a different rate) for $\hbar \omega < E_g$. The oscillatory behaviour is known as the Franz-Keldysh Effect, and the oscillations themselves referred to as Franz-Keldysh Oscillations (FKOs). The physical interpretation of this lineshape involves the spatial separation of the electron and hole wavefunctions due to the in-built electric field as shown in Figure 2.4, causing them to collect at opposite ends of the tilted band profiles. For $\hbar \omega < E_g$, only the portions of the wavefunctions which have tunnelled into the potential barrier can overlap, giving rise to the evanescent tail. For $\hbar \omega > E_g$, overlap of the oscillatory portion of their wavefunctions is possible as normal. However, due to the electric field those wavefunctions have slightly different phases, resulting in an interference pattern which manifests as the oscillatory component of the PM lineshape. If $\hbar \omega$ is large enough that the energy of the transitioning electrons is greater than the potential energy of the field, the field’s effect on them is negligible. The joint density of states at those energies thus remains unaltered, and optical transitions proceed as in the zero-field case (i.e. momentum conserving transitions only). The PM signal must, therefore, die away with increasing energy, leading to the decay of the FKO.

Figure 2.4 Schematic illustrating the origin of FKO in a bulk semiconductor with a non-zero internal electric field. Electrons and holes move in opposite directions under the field's influence (LHS). The wavefunctions of electrons, $\Psi_e$, and holes, $\Psi_h$, overlap within the bandgap (i.e. $\hbar \omega < E_g$) due to tunnelling effects. Resulting optical transitions lead to a non-zero PM signal (RHS) at $\hbar \omega < E_g$. A field-induced phase difference in the wavefunctions at $\hbar \omega > E_g$ leads to exponentially decaying oscillations, known as Franz-Keldysh Oscillations (FKOs).
An additional element one must consider is the effect of carrier-carrier and carrier-lattice scattering processes, which result in a finite average lifetime, $1/\tau$, for an electron in a Bloch state. Through the uncertainty principle (i.e. $\tau E \geq \hbar/2$), such interactions tend to smear out the energies of Bloch states in an effect referred to as lifetime broadening. This phenomenon is incorporated into modulation spectroscopy theory by making $F(\xi_0)$ and $G(\xi_0)$ functions of the following complex variable:

$$\xi_0 = \frac{E_g - i\hbar\omega}{\hbar\Omega} + i \frac{\Gamma}{\hbar\Omega}$$

(2.26)

where $\Gamma$ is the lifetime broadening parameter. By increasing the ratio of $\Gamma$ to $\hbar\Omega$ one can therefore damp the FKO oscillations produced by $F(\xi_0^*)$ and $G(\xi_0^*)$ in a manner akin to placing a mass on the end of a spring. Figure 2.5 shows an example of a theoretical lineshape one might expect a practical PM signal (including the effects of lifetime broadening) to approximate.

**Figure 2.5** Schematic of a typical (theoretical) lineshape that might result from an experiment which modulates the internal field of a bulk semiconductor in the high-field regime. In this instance $\hbar\Omega=35\text{meV}$, $\Gamma=15\text{meV}$, and the phase angle (i.e. determining the relative contributions of $\varepsilon_1$ and $\varepsilon_2$) is $45^\circ$. $E_g$ is indicated. Of particular note is the non-zero amplitude for $\hbar\omega<E_g$, and the oscillatory behaviour for $\hbar\omega>E_g$. The latter are referred to as Franz-Keldysh Oscillations (or FKOs).

If the magnitude of the energy gained by an electron in the electric field $e\mathbf{E}z$ is small
compared to \( \Gamma \) (i.e. \( \hbar \Theta << \Gamma \)), the imaginary part of equation 2.26 is very large, and the FKO oscillations are heavily damped. In what is known as the low-field regime, a simple argument can then be made which that PM signal lineshapes reduce to a simpler third derivative-like form. This argument goes as follows: The force on the electron in the field is \( eE = \mu \delta v / \delta t \), where \( \delta v / \delta t = eE / \mu = \delta^2 z / \delta t^2 \). Thus \( z = (eE^2) / 2\mu \), and \( e\delta z = (eE^2) / 2\mu \). The electric field induced change in \( \varepsilon \), \( \Delta \varepsilon = \varepsilon (E_\psi + e\delta z) - \varepsilon (E_\psi) \), may then be expressed as a Taylor power series. Assuming only a small perturbation the first order term of the series is sufficient:[26]

\[
\Delta \varepsilon (E_\psi, \xi) \approx \frac{e^2 \xi^2 t^2}{2\mu} \frac{\partial \varepsilon}{\partial E_\psi} \quad (2.27)
\]

The quantum mechanical operator for \( t \) is \( i\hbar d/dE_\psi \), so \( \Delta \varepsilon \) is thus shown to have the following third derivative form:[26]

\[
\Delta \varepsilon (E_\psi, \xi) \approx (\hbar \Theta)^3 \frac{\partial^3 \varepsilon}{\partial E_\psi^3} \quad (2.28)
\]

where the dependence of \( \Delta \varepsilon \) on the square of the field magnitude may be noted. A seminal work in modulation spectroscopy theory, in particular those concerned with internal electric field modulation, is that of Aspnes.[21] There, an analytical expression is derived which approximates the third derivative form with a Lorentzian-like function, including in its definition \( \Gamma \) and the Seraphin coefficients (equation 2.17). This generalised ‘third derivative functional form’ (or TDFF) at a parabolic (or \( M_0 \)) critical point reads as follows:

\[
\frac{\Delta R}{R} = \text{Re} \left[ Ce^{i\theta} \left( \hbar \omega - E_g + i\Gamma \right)^{-n} \right] \quad (2.29)
\]

where the \( Ce^{i\theta} \) is the now familiar representation of the presumed unknown Seraphin coefficients, where \( C \) now includes the electro-optic energy, and \( \theta \) defines the relative weights of \( \varepsilon_1 \) and \( \varepsilon_2 \). The exponent \( n \) is a phenomenological variable related to the dimensionality of the critical point. For a 1D critical point \( n=3.5 \), for 2D \( n=3 \), and for
3D $n=2.5$. Excitons have zero dimensionality (i.e. a singularity on the scale of semiconductor heterostructures) and may be represented by $n=2$. [21] In Table 9.1 (see appendix, section 9.4) it is shown that all critical point types exhibit similar $E^{1/2}$-like energy dependences, so one can reasonably assume that equation 2.29 also provides a reasonable approximation to critical point types other than $M_0$ also. Figure 2.6 shows an example of a theoretical TDFF lineshape. There it can be seen that the relevant critical point energy ($E_g$ in this case) lies close to the centre of the feature. This is a general result for TDFF lineshapes, and will be used extensively as an approximation during the analysis of many of the PM results presented in this thesis.

![Figure 2.6 Schematic of a typical (theoretical) lineshape that might result from an experiment which modulates the internal field of a bulk semiconductor in the low-field regime, as approximated by the third derivative functional form (equation 2.29). Here, $\Gamma=15\text{meV}$, $\Theta=45^\circ$, and $n=2.5$ (i.e. a 3D $M_0$ critical point). Note the proximity of $E_g$ to the centre of the lineshape. For a symmetric lineshape (i.e. $\Theta=0^\circ$), $E_g$ would coincide precisely with the lineshapes mid-point. No FKOs are observed in this regime/approximation.](image)

To summarise, when $h\Theta>>\Gamma$ (i.e. the high-field regime) extended, FKO-like PM features are observed, while conversely if $h\Theta<<\Gamma$ (i.e. the low-field regime) the signal reduces to a more localised third derivative form with only two or three lobes. There is a smooth transition between these two extremes, and the lineshape in the resulting 'intermediate-field' regime appears intermediate between a TDFF-like and FKO-like lineshape; in essence a damped FKO. It is worth pointing out that although it is conventional to refer to the field strength as defining these regimes, it is equally valid for the so-called high- or intermediate-field regimes to be encountered if $\Gamma$ itself is
very large (i.e. in the presence of strong scattering). Indeed, in section 3.3.5 PM results on InSb exhibiting an anomalous temperature dependence are presented, which it is believed may be explained in terms of changes in $\Gamma$, rather than $\hbar\Theta$.

The discussion presented thus far applies equally to any experimental technique which modulates the internal field of a sample. In electroreflectance (ER) experiments, applied internal fields of the order kV/cm can be readily achieved, and the high-field regime is frequently encountered. However, PM signals are entirely dependent on the magnitude of a sample’s in-built electric field. In a bulk semiconductor, any such field is likely to be significantly smaller than any internal field attainable by applied means. Therefore, the observed lineshapes in such measurements are usually limited to the low- and intermediate-field regimes, exhibiting only weak FKOs if any.

The PT mode of PM spectroscopy works via precisely the same mechanism as PR, only the Seraphin coefficients must now take the following form:

$$\alpha_T = \frac{1}{T} \frac{\partial T}{\partial \varepsilon_1} \quad \text{and} \quad \beta_T = \frac{1}{T} \frac{\partial T}{\partial \varepsilon_2} \quad (2.30)$$

where Fresnel’s equation for transmittance, $T$, for normal incidence on a semiconductor-air interface can be expressed in terms of $\varepsilon_1$ and $\varepsilon_2$ as follows:[18]

$$T = \frac{4\left[(\varepsilon_1^2 + \varepsilon_2^2)^{1/2}\right]^{1/2}}{\left[(\varepsilon_1^2 + \varepsilon_2^2)\right]^{1/2} + \left[2\varepsilon_1 + 2\left(\varepsilon_1^2 + \varepsilon_2^2\right)^{1/2}\right]^{1/2} + 1} \quad (2.40)$$

This is merely the transmittance equivalent of equation 2.14. Because the only difference between reflectance and transmittance mode is a modification of the Seraphin coefficients, the theoretical lineshapes as given by equation 2.29 should be similar in both cases, save for changes in $C$ and $\theta$. It should be noted, however, that absorption within the sample can have a far greater effect on a PT lineshape than in PR, potentially affecting both its amplitude and phase. For example, if $T \to 0$ within
the range of a PT feature (i.e. a sharp absorption edge or cut-off), premature truncation of the lineshape can occur.

### 2.3.2 Practical Implementation

A typical PR experiment consists of a mechanically chopped laser beam (or pump) incident on the surface of the semiconductor sample of interest, arranged such that it overlaps a spectrally resolved broad band beam (or probe)\(^4\). The reflected component of the probe is then detected via a lock-in amplifier, the output of which is the difference between laser on and laser off states, or \(\Delta R\). A true PR signal requires both \(R\) and \(\Delta R\) (i.e. \(\Delta R/R\)), so it is common to split the signal before the lock-in, thus allowing the dc component (\(R\)) to be measured simultaneously with the ac component (\(\Delta R\)), perhaps via a multimeter. This minimises any intensity drift in the pump or probe which might appear in otherwise separate \(R\) and \(\Delta R\) measurements. Typical low-field regime \(\Delta R/R\) signals are of the order \(10^{-5}-10^{-3}\). The key components of a typical MIR PR set up are illustrated in Figure 2.7. The arrangement for PT is fundamentally similar, only in that case it is the transmitted light which is collected and passed through the monochromator.

The photon energy of the modulating laser is generally required to be larger than the bandgap of the material of interest. Typically the signal at the detector is a mixture of scattered laser light, PL from the sample, and \(\Delta R\). All three of these signals are modulated at the laser chopping frequency, so all three may appear in the lock-in output. The former two components give rise to a wavelength independent offset upon which \(\Delta R\) sits. As long as this background is stable, it can be easily removed during data analysis. However, it is often the case that over the period of acquisition the \(\Delta R\) background is seen to vary with time, mainly due to laser output drift and/or fluctuations. Great care must then be taken to minimise the amount of laser scatter reaching the detector. This is usually achieved by adding a high pass optical filter before the detector, chosen such that it passes the probe wavelengths of interest, but

\(^4\)If the probe beam’s angle of incidence is not 0° (i.e. normal incidence), the Seraphin coefficients become considerably more complicated and harder to manage mathematically, but ultimately all that does is change \(C\) and \(\theta\).
blocks those of the laser line(s). Other precautions including sensible placement of beam stops and beam paths can also help. Unfortunately optical filters cannot be used to screen out the PL component of the signal, as it is likely to overlap in wavelength with the $\Delta R$ features. However, several variations on the basic arrangement depicted in Figure 2.7, designed to remove the PL component during acquisition, are available. These techniques tend not to be universally applicable though, and must be implemented on a case-by-case basis. Often the PL intensity is insignificant compared to the laser scatter signal anyway and, therefore, is not always of primary concern. This should be particularly so in bulk NGSs, where non-radiative recombination mechanisms are orders of magnitude more efficient than those in WGSs.

![Figure 2.7 Schematic of a typical MIR PR arrangement.](image)

It has been found that chopping frequency can have a dramatic effect on the size of PR signals. This has been explained in terms of trap states, which can capture

---

5 For this reason, the modulation laser photon energy is usually chosen to be sufficiently far above the bandgap that there is no chance of overlap with anticipated $\Delta R$ features.
Optical Characterisation Techniques

photo-excited carriers and hold them on time scales of the order of milliseconds. These trapped carriers continue to screen the in-built field for this duration. If the period of the laser modulation is shorter than the trapping time, the in-built field remains in a strongly screened state invariant with time, thereby reducing the amplitude of $\Delta R$. Reductions in $\Delta R/R$ as large as a factor of $\sim 4$ have been observed in GaAs for chopping frequencies of the order of $\sim 1$kHz, with the signal continuing to decrease steadily for still higher frequencies. Therefore, modulation frequencies are usually kept well below $\sim 1$kHz so that any such effects are minimised.

2.4 References


3 First Application of FTIR Spectroscopy to Photomodulation Techniques

3.1 Introduction

As was stated previously, FTIR spectrometers, particularly those based on the Michelson design, are considered the instrument of choice for MIR applications (e.g. PL and absorption measurements). However, to the author's knowledge, prior to this work there had not yet been a successful demonstration of PM spectroscopy performed using FTIR techniques, with the longest wavelength PM measurement standing then at a wavelength corresponding to the room temperature InAs bandgap (i.e. ~3.5μm). That measurement was performed using a grating spectrometer-based system.[1] Believing that a resistance to moving to FTIR-based systems for MIR PM applications presents a major barrier to extending such techniques to longer wavelengths, the aim of this work was to perform the first proof of principle measurements demonstrating ‘Fourier transform photomodulation spectroscopy’ (or ‘FTPM’ as it has been coined).

3.1.1 Choosing a Suitable FTIR System

As alluded to in section 1.1.3 previously, Michelson FTIR spectrometers come in two distinct designs; continuous scanning and step-scanning. The former may be further classified as either fast-scanning or slow-scanning. Spectrometers of this type are so called because the moving mirror is scanned at some continuous velocity through a given retardation range. This provides two important advantages over step-scanning;
firstly the faster the mirror is moved over a given retardation range the more rapidly scans can be acquired, and secondly the frequency at which the fringes are modulated by the mirror's movement can be used to apply a form of phase-sensitive detection without the need for mechanical chopping. Consequently, fast-scanning FTIR spectrometers are less sensitive to slow variations in source intensity, while also affording a factor two increase in collection efficiency (i.e. without chopping twice as many photons can be collected on average). However, as will now be explained, it is the resulting rapid fringe modulation that makes such advantages possible which also represents the biggest obstacle to adapting FTIR spectroscopy for PM.

Consider a typical commercial fast-scanning Michelson FTIR system. Alignment and calibration are performed using an internal 632.8nm HeNe laser, engineered to follow the same path as the measured MIR light. After the beamsplitter this laser is picked off and directed onto a photo-diode where the resulting interferogram is detected. Since the laser is an extremely narrow-band source, the interferogram has a long coherence length, with close to perfectly dark and perfectly light alternating fringes of periodic length $\lambda_L=632.8\text{nm}$ (see appendix, section 9.2). For a mirror moving with a continuous velocity $v \sim 2\text{mms}^{-1}$, the fringes pass across the photo-diode at a frequency, $f_m^L$, given by:[2]

$$f_m^L = \frac{2v}{\lambda_L} \quad (3.1)$$

An ac signal is thus detected at the photo-diode, and the fringe intensity is said to be modulated at frequency $f_m^L$ (typically in the audio range). Continuous scanning systems use this fringe pattern to track the mirror's position and synchronise data sampling (i.e. one period of the pattern corresponds to a mirror movement of $\lambda_L/2$). All the component wavelengths of the MIR light comprising the interferogram of interest must also be modulated in this way. If the shortest of these wavelengths is, for example, $\lambda_{\text{min}} \sim 1.5\mu\text{m}$ then for the same mirror velocity the maximum modulation frequency, $f_m^{\text{max}}$, is given by:

---

6 The intensity ratio of the light and dark fringes can also be used to optimise the system's alignment (e.g. optimum alignment gives the largest ratio)
The MIR detector signal will, therefore, contain a broad band of ac components up to and including \( -2.666 \text{kHz} \). These may confuse any phase sensitive treatment of signals external if the external modulation frequency (e.g. a mechanical chopping frequency) is smaller than or comparable to \( f_{m}^{\text{max}} \). In order to avoid this problem, the external modulation must be performed at a frequency several times larger than \( f_{m}^{\text{max}} \), preferably at least five times.[2] Since PM signals are generally found to reduce quite rapidly at modulation frequencies higher than \(~1\text{kHz}\), applying a chopping frequency as large as \( 13.3\text{kHz} \) (in this example) is clearly unworkable.[3,4] In particular, the fast-scanning interferometer available at Surrey has a maximum mirror modulation frequency of \(~3\text{kHz}\), leading to a minimum chopping frequency of at least \( 15\text{kHz} \). For this reason, the fast-scanning FTIR spectrometer available at Surrey could not be used for PR measurements.

Instead it was decided a step-scanning FTIR spectrometer offered the best chance of successfully demonstrating an FTPM technique. However, purchasing a brand new commercial step-scanning interferometer was not practical for this project. Instead it was chosen to adapt a ‘retired’ system for the task. In fact, such practicalities aside, this approach is ultimately advantageous at this early stage of development, as commercial step-scanning systems often employ phase sensitive detection internally to improve the SNR (e.g. mechanical chopping, mirror vibration/modulation, etc.).[2] Depending on the frequency that modulation occurs at, even step-scanning systems can thereby be rendered unsuitable for PM applications, and for precisely the same reasons as fast-scanning models. By retro-fitting an older system as required, such problems can be eliminated during the design phase.

While the bulk of the work performed during this project was development of the step-scanning FTIR system and its integration into an FTPM arrangement, much of it is engineering in nature and will not be described here. For a detailed account of that work, the interested reader is directed to the appendix of this thesis (section 9.3).
3.2 Theory

3.2.1 Interferometric PM Signals

Prior to attempting FTPM measurements, an investigation was performed into the kind of signals one might expect to find. This involved generating some theoretical TDFF lineshapes (i.e. signals in low field regime) as described by equation 2.29, and Fourier transforming them to obtain their interferometric equivalents. Figure 3.1 summarises this process, illustrating also the relationship between the modulation signals in each domain and the perturbed (laser on) and unperturbed (laser off) spectra/interferograms. Note reflectance mode is assumed in this analysis purely for the sake of argument. The theoretical $\Delta R$ lineshape in Figure 3.1 was arbitrarily chosen to have zero phase (i.e. $\theta=0$ in equation 2.29). In Figure 3.2 the effect of the lineshape phase on the interferogram is shown. There it is seen that the envelope of the interferogram remains unchanged, with only the phase of the fringes changing by $\theta$. Note that this is not the same as adding a phase error to the interferogram (see appendix, an effect described in the appendix, section 9.2.4).

A spectrum containing just one TDFF-like $\Delta R$ feature is, therefore, expected to result in interferometric signals like those illustrated in Figure 3.2. The absence of an obvious centreburst in such signals implies that the spectral integral of a TDFF lineshape is zero, which is indeed so. Of course the usual correlations between spectral width and interferogram coherence length apply to $\Delta R$ signals (see appendix, section 9.2), so a narrow $\Delta R$ lineshape yields a broad interferogram and vice versa. Generally, a spectrum might contain more than one $\Delta R$ feature, corresponding to multiple critical points. In that case the resulting interferogram would exhibit a beat pattern whose period represents the separation of the features in the frequency domain. Furthermore, in the intermediate-field regime $\Delta R$ will, in general, deviate from the TDFF form itself. Hence, a practical measurement in that regime could (in principle) yield almost any arbitrary interferometric signal, with Fourier components that depend on the period and rate of decay of any FKOs.
Figure 3.1 TDFF-like (i.e. low-field) $\Delta R$ signals in the frequency ($\text{cm}^{-1}$) and real space (cm) domains. On the LHS, the reflectivity spectra with modulation laser off, $R_{\text{spe}}(\text{off})$, and laser on, $R_{\text{spe}}(\text{on})$, are shown (both are black body distributions for the sake of argument). It is the difference between those spectra, $R_{\text{spe}}(\text{off})-R_{\text{spe}}(\text{on})$, that one measures as $\Delta R_{\text{spe}}$. Note that $\Delta R_{\text{spe}}/R_{\text{spe}}\approx 10^{-4}$ typically, so $\Delta R_{\text{spe}}$ is too small to be discerned in $R_{\text{spe}}(\text{off})$, although its position is indicated. The corresponding reflectivity interferograms (related to the spectra by an inverse FT, and vice versa) are shown on the RHS. $\Delta R_{\text{igram}}$ is then the difference between those interferograms; $\Delta R_{\text{igram}}=R_{\text{igram}}(\text{off})-R_{\text{igram}}(\text{on})$. Note that $\Delta R_{\text{spe}}$ and $\Delta R_{\text{igram}}$ are related via a Fourier transform pair, just like the reflectivity spectra. $\Delta R_{\text{igram}}/R_{\text{igram}}$ is too small for $\Delta R_{\text{igram}}$ to be discerned in $R_{\text{gram}}(\text{off})$ also.
3.3 Experiment

In order to gauge the quality/validity of the FTPM results, where possible they have been compared with those obtained on the tried and tested grating system. Three samples were chosen for this comparison; an InGaAs quantum well structure, bulk GaSb, and bulk InAs. Each presents a very different material system or structure (and wavelength range) for investigation, and all were found to give strong PR signals measured using the grating spectrometer. As a final step in this proof of principle, temperature dependent PT measurements have been performed on two nominally
similar InSb samples in an effort to move FTPM beyond the limits of the grating system (so no corroborating measurements using the grating set up were possible), into its optimum wavelength region.

### 3.3.1 The Step-Scan FTPM Arrangement

The FTPM optical arrangement is shown in Figure 3.3. In this instance the PT mode of operation is illustrated, although the PR setup is identical until the beams reach the sample. The arrangement is presented here mainly for ease of reference. For a more detailed description of its components the reader should refer to the appendix of this thesis (section 9.3.2). Aside from several minor peculiarities detailed therein, the arrangement is largely consistent with the generic MIR PR arrangement described previously in section 2.3.2.

![Figure 3.3](image)

**Figure 3.3** A schematic of the FTPM (or more specifically FTPT) optical arrangement. For a detailed description of the individual components refer to the appendix of this thesis (section 9.3.2).
3.3.2 InGaAs/InGaAsP Quantum Well

This sample consisted of four 950nm wide In$_{0.533}$Ga$_{0.477}$As quantum wells with 1400nm wide In$_{0.80}$Ga$_{0.20}$As$_{0.43}$P$_{0.57}$ barriers. The barrier layers were designed to be lattice matched to an InP substrate. A grating-based PR study of this sample has been performed previously.[5] The lowest energy transition in the well is that between the conduction band and heavy hole ground states, H$_{11}$, observed at a wavelength of ~1580nm (0.784eV). A range of higher energy transitions between confined states were also observed down to ~1300nm (0.935eV), with a barrier-related FKO at around 1180nm (1.053eV). The wavelengths of the PR features in this sample approach the short wavelength limit of the step-scanning FTIR spectrometer’s operational range, making this measurement a demanding test of the FTIR spectrometer’s capabilities.

3.3.2.1 Experimental Arrangement

This measurement was re-done using the FTPR arrangement at room temperature, so no cryostat was required. Otherwise the experimental arrangement was as shown in Figure 3.3 (only in PR mode rather than PT). A 10mW HeNe laser was used as the modulation source, mechanically chopped at ~800Hz. The amplified detector output contained both dc and ac components, each being passed directly to a multimeter and lock-in amplifier respectively. The lowest permissible mirror step-size of 1×10$^{-5}$cm (equivalent to an optical retardation of 2×10$^{-5}$cm) was used, yielding a Nyquist limit of ~400nm (3.108eV) (see appendix, section 9.2.2). The tungsten-halogen source and quartz glass beamsplitter were the best choice for optimal light throughput in this NIR wavelength range, with R and ΔR detected using an InGaAs photo-diode with long wavelength cut-off at ~1620-1650nm (~0.76-0.75eV). A 1200nm (1.04eV) long-pass filter was placed at the detector position to block short wavelengths. Much of the tungsten-halogen source light occurs at wavelengths between 400nm (3.11eV) and 1000nm (1.24eV), so by blocking those wavelengths an improvement in SNR (due to the elimination of their contribution to vibrational noise within the system) was achieved.

A single scan of range 1.75×10$^{-2}$cm (centred on ZPD) was performed, leading to a true resolution of ~28.6cm$^{-1}$ (~3.5meV). Subsequent zero-filling of the interferogram
(see appendix, section 9.2.2) for input into the FFT function resulted in an apparent resolution of \(~24\text{cm}^{-1}\) (\(~3.0\text{meV}\) between \(~0\text{cm}^{-1}\) and \(25000\text{cm}^{-1}\).

### 3.3.2.2 Results and Discussion

The spectra measured using the grating\(^7\) and FTPR systems are compared in Figure 3.4. Above \(~1250\text{nm}\) the two give results in good agreement with one another, with all the features observed in the grating result identifiable in the FTPR spectrum also. These six transitions are labelled in accordance with their previous assignments.[5] The FTPR spectrum is truncated by the \(1200\text{nm}\) long-pass filter, so as one might expect the one remaining quantum well transition and barrier related FKO at shorter wavelengths are missing from the FTPR result. However, it may be noted that the barrier feature has been observed in lower quality FTPR scans (not shown), recorded without the long-pass filter in position.

Figure 3.4: Comparison of an InGaAs/InGaAsP quantum well sample's room temperature \(\Delta R\) spectrum, measured using the grating (blue) and FTIR (red) PR systems. The inset shows the processed interferogram (i.e. apodised, phase corrected, zero-filled, etc. – see appendix, section 9.2) corresponding to the FTIR spectrum. The spectrum itself has been scaled and shifted downwards for clarity. All transitions at wavelengths longer than \(1250\text{nm} (~1\text{eV})\) are present in both spectra. At shorter wavelengths, the FTIR spectrum is truncated due to the use of a \(1200\text{nm}\) long-pass filter. Note the region \(1210\text{nm}-1510\text{nm}\) is magnified by a factor of three. Transition assignments are those determined from the previously published grating result on this sample.[5]

\(^7\) The grating result presented here is reproduced with the permission of the author (Dr. T. J. C. Hosea).[5]
There is a slight difference in wavelength apparent between the two spectra in Figure 3.4. This shift is most noticeable at the transition energies, where all the lineshapes in the FTPR spectrum appear blueshifted by ~5nm relative to their grating versions. As a phase corrected interferogram always has one unique point of reference in ZPD, only a consistent positional inaccuracy in the translation stage's movement (relative to ZPD) could produce such a shift in an FTPR spectrum. Any positional error is more likely to be random in nature, so it is felt the error is unlikely to lie with the FTIR system. Rather this is most probably a calibration error in the grating system.

### 3.3.3 Bulk GaSb

This sample is a piece of n-type GaSb substrate ($n = 2 \times 10^{17} \text{cm}^{-3}$), provided by Lancaster University. It exhibits a strong room temperature PR signal, so no cryostat was required. The stated doping level should shift the absorption edge of the sample from the intrinsic value $E_0(T=295K) \approx 1700\text{nm}$ to ~1590nm (~0.73eV to ~0.78eV).[6]

The grating result was in this case measured using the lower specification of the two spectrometer models described in section 1.1.2.3. These were carried out by Dr. T. J. C. Hosea.

### 3.3.3.1 Experimental Arrangement

The FTPR experiment was performed in much the same way as for the previous sample. The same 10mW HeNe laser was used as the modulation source, this time chopped at ~200Hz. Unfortunately, the GaSb bandgap lies extremely close to the InGaAs detector long wavelength cut-off, preventing its use on this occasion. Instead a nitrogen-cooled InSb detector was used to measure $R$ and $\Delta R$. This detector's matched pre-amplifier does not pass dc signals, so to get around this problem an alternative Stanford amplifier was employed, with a 3kHz low-pass filter selected. That instrument's output was then distributed to the multimeter and lock-in amplifier as normal. An 850nm long-pass filter was also used to help block laser scatter at the detector position, although the InSb detector's sensitivity at wavelengths shorter than ~1000nm is already relatively low to begin with (<5% peak sensitivity).
A mirror step-size of $2 \times 10^5 \text{cm}$ was used, scanning over a range $5 \times 10^3 \text{cm}$ wide, leading to a true resolution of $\sim 100 \text{cm}^{-1}$ ($\sim 12.4 \text{meV}$). Subsequent zero-filling of the interferogram for input into the FFT function resulted in an apparent resolution of $\sim 98 \text{cm}^{-1}$ ($\sim 12.2 \text{meV}$) between $0 \text{cm}^{-1}$ to $12500 \text{cm}^{-1}$. A total of 19 individual scans were averaged to produce the final spectrum.

### 3.3.3.2 Results and Discussion

The grating and FTPR results on this sample are compared in Figure 3.5. As with the quantum well sample earlier, there is good agreement between the two spectra at wavelengths longer than $\sim 1200 \text{nm}$. In this case the only feature at those wavelengths is that associated with the bulk fundamental transition, $E_0$. The features observed in the grating spectrum at $\sim 830 \text{nm}$ ($\sim 1.49 \text{eV}$, $E_0 + \Delta_0$), $\sim 600 \text{nm}$ ($\sim 2.07 \text{eV}$, $E_1$), and $\sim 490 \text{nm}$ ($\sim 2.53 \text{eV}$, $E_1 + \Delta_1$) are consistent with the results of Hwang et al, though their wavelengths are far too short for accurate measure by the FTPR system.[7]

![Figure 3.5](image_url)  
**Figure 3.5** Comparison of bulk $n$-type GaSb $\Delta R$ spectrum measured at room temperature using the grating (blue) and FTIR (red) PR systems. Shown inset is the processed interferogram corresponding to the FTIR spectrum. The spectrum itself has been scaled and shifted downwards for clarity. Only the bulk bandgap feature, $E_0$, lies at long enough wavelength to be reliably measured by the FTPM system, although the resulting lineshape in that case is seen to closely follow the grating result. The expected wavelength corresponding to $E_0$, taking account of the anticipated Moss Burstein shift in absorption due to doping ($n \sim 2 \times 10^{17} \text{cm}^{-3}$), is indicated.[6] Feature assignments match those given by.[7]
The $E_0$ feature in Figure 3.5 is around 1000nm in extent, about thirty times broader than the dominant ground state transition, $H_{11}$, in the InGaAs quantum well sample (see Figure 3.4). Thus the interferogram (shown inset in Figure 3.5) is also around thirty times narrower here than in that earlier case. The high energy side of the $E_0$ lineshape shows additional oscillations consistent with FKOs.

### 3.3.4 Bulk InAs

This sample, also provided by Lancaster University, is a piece of $p$-type InAs substrate ($p \sim 1 \times 10^{17} \text{cm}^{-3}$). The room temperature signal from this sample was extremely weak even in the grating set up, so the comparative measurements were performed at $77K^8$. The $p$-type doping is expected to shift the absorption edge of the sample from its intrinsic value $E_0(T=77K) \approx 3050\text{nm}$ to $\sim 2980\text{nm}$ ($-0.406\text{eV}$ to $-0.416\text{eV}$).

The grating result was again measured using the lower specification of the two spectrometer models described in section 1.1.3.3, by Dr. T. J. C. Hosea.

#### 3.3.4.1 Experimental Arrangement

The experimental procedure was similar to that described previously, only for this sample cryogenic temperatures are required to obtain a strong signal. Therefore, the sample was mounted on the cold finger of an Oxford Instruments nitrogen bath cryostat. Thermal contact was made via a copper plate, to which the sample was attached using silver conductive paint. The graphite globar source and KBr beamsplitter combination was the optimal choice for throughput at $\sim 3\mu\text{m}$. A 10mW HeNe laser was again used as the modulation source with chopping frequency $\sim 200\text{Hz}$. The nitrogen-cooled InSb detector was used as described previously (see section 3.3.3.1). Unfortunately the substrates of the available long-pass optical filters are opaque between $\sim 2500-3000\text{nm}$ and above $\sim 3500\text{nm}$, prohibiting their use in screening scattered laser light in this case.

---

8 It was hoped that the smaller carrier density at low temperature would increase the inbuilt field due to decreased screening by residual carriers. The PR signal was indeed found to increase with cooling.
A mirror step-size of $3 \times 10^{-5}$ cm was used, scanning over a range $6 \times 10^{-3}$ cm wide, leading to a true resolution of $\sim 83 \text{cm}^{-1}$ ($\sim 10.3 \text{meV}$). Subsequent zero-filling of the interferogram in preparation for FFT resulted in an apparent resolution of $65 \text{cm}^{-1}$ ($\sim 8.1 \text{meV}$) between $0 \text{cm}^{-1}$ to $8268 \text{cm}^{-1}$. A total of 8 individual scans were averaged to produce the final spectrum.

### 3.3.4.2 Results and Discussion

The grating and FTPR results on this sample are shown in Figure 3.6. There is obvious similarity between the $E_0$ feature lineshape in the measurements as hoped. However, the two are not as alike as was found for the $n$-GaSb and InGaAs QW samples. The spin-orbit feature, $E_0 + \Delta_0$, at $\sim 1540 \text{nm}$ ($\sim 1.24 \text{eV}$) is both too weak and at too short a wavelength for detection by the FTIR when in this MIR globar/KBr beamsplitter configuration (lower limit $\sim 2 \mu$m). However, if the FTIR were optimised for NIR operation (i.e. quartz source and beamsplitter) there is no reason (in principle) why the $E_0 + \Delta_0$ feature should not also be observed.

![Figure 3.6 Comparison of bulk $p$-type InAs PR spectrum, measured at nitrogen temperature (77K) using the grating (blue) and FTIR (red) PR systems. Shown inset is the processed interferogram corresponding to the FTIR spectrum. The spectrum itself has been scaled and shifted downwards for clarity. Only the bandgap feature, $E_0$, lies at long enough wavelength to be reliably measured by the FTPR system when in its MIR configuration. The two lineshapes are seen to be broadly similar, though there is significant deviation at long wavelength. The wavelength corresponding to $E_0$, taking account of the anticipated Moss Burstein shift in absorption due to doping ($p = 1 \times 10^{17} \text{cm}^{-3}$), is indicated.]
There is a clear discrepancy between the two results on the long wavelength side of the $E_0$ feature, with the grating result exhibiting an oscillatory shape which rapidly decays to zero, while the FTPR result shows an approximately constant non-zero tail. The grating spectrometer’s upper wavelength limit was $\sim 3\mu$m, however, so it is likely $\Delta R/R$ beyond that point is adversely affected by the $R$ signal dying away to zero. It is, therefore, believed the FTPR result is the more accurate at long wavelength. Evidence corroborating this statement is provided by measurements made some time later using the larger grating system (and He closed-cycle cryostat), the operational wavelength range of which extends to $\sim 4500$nm (0.28eV). Such a result is not used for comparison here, however, as no 77K spectrum was available. It is not clear why the PR signal does not decay to the baseline value at long wavelength. The predicted position of $E_0$, as indicated in Figure 3.6, is close to the lineshape’s centre, as expected for a TDFF-like feature’s critical point energy (see section 2.3.1). Fitting with a TDFF lineshape is likely to be difficult, however, due to the feature’s long non-zero tails.

### 3.3.5 Bulk InSb

Two nominally intrinsic bulk InSb samples, grown for an unrelated project by Qinetiq have been measured using the FTPT technique. Both consisted of $\sim 2\mu$m thick epilayers grown on GaAs substrates. A list of the available sample details are provided in Table 3.1.

<table>
<thead>
<tr>
<th>Sample Code</th>
<th>Structure</th>
<th>Carrier Concentration (cm$^{-3}$) [type]</th>
<th>Mobility (cm$^2$V$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME1681</td>
<td>GaAs/2.25μm InSb</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ME1655</td>
<td>GaAs/InSb</td>
<td>$2.8 \times 10^{15}$ [n] at 77K</td>
<td>2.99$\times 10^4$ at 77K</td>
</tr>
</tbody>
</table>

*Table 3.1 Details of the nominally intrinsic bulk InSb samples used for this FTPT study.*

Literature sources suggest the intrinsic carrier concentration and mobility at 77K for InSb grown on GaAs substrate are $\sim 2.2 \times 10^{15}$cm$^{-3}$ and $\sim 1.5 \times 10^5$cm$^2$V$^{-1}$s$^{-1}$ respectively. Hence, although the carrier concentration of ME1655 is consistent with a low (unintentional) doping level, its mobility is five times smaller than the corresponding
literature value. Such a small mobility can be explained, however, if one notes that the lattice constants of GaAs and InSb are 5.66Å and 6.48Å respectively, corresponding to a ~13% mismatch. Therefore, it is likely that the large compressive strain introduced on the InSb side of the substrate/epilayer interface has resulted in large numbers of dislocations, reducing the mobility accordingly. Furthermore, the bandgap of ME1655 might also be larger than that of pure InSb (suggested value $E_g \approx 0.173\text{eV}$, at room temperature) due to the reduction in lattice constant induced by the strain.[9] Though no precise data are available for ME1681, it is expected that this sample possesses broadly similar properties to ME1655.

### 3.3.5.1 Experimental Arrangement

The experimental procedure was similar to that described previously, with cryogenic temperatures again expected to be necessary to obtain a strong signal (as with InAs). However, on this occasion the aim was to perform a temperature dependent study. To that end, the sample was mounted on the cold finger of the nitrogen bath cryostat as detailed earlier. The graphite globar source and KBr beamsplitter combination were the only viable choice for operation beyond 5μm. Modulation was provided by a diode-pumped Nd:YAG laser, outputting ~100mW at a wavelength of ~1064nm under optimal conditions, with mechanical chopping performed at a frequency of ~800Hz. Detection at these long wavelengths was achieved using a nitrogen-cooled HgCdTe photo-detector. As with the InSb detector utilised for previous measurements, the primary output of its matched pre-amplifier filtered out dc signals. However, the solution employed in that case could not be used here, since the Stanford filter/amplifier was unable to provide sufficient gain to turn the HgCdTe detector’s output into a measurable voltage signal at the lock-in amplifier. Instead a secondary output on the HgCdTe detector’s matched pre-amplifier was used for the dc. This provides a weaker, unfiltered signal, and is intended only as a test point for the pre-amplifier’s internal circuitry. The signals at this point are, therefore, relatively small and noisy, but nevertheless found to be adequate for measurement of the dc signal level. It should be noted that taking the dc and ac signals from outputs with different (and unknown) gain values in this way makes the absolute values of $\Delta T/T$ meaningless. There were no available high-pass optical filters available suitable for this wavelength range.
The FTIR mirror was stepped in increments of $5 \times 10^{-5}$ cm. The scan range was varied from temperature to temperature as the width of the lineshape evolved, though when it was at its narrowest (at 77K) the largest range was $5 \times 10^{-2}$ cm, leading to a true resolution of $\sim 10 \text{cm}^{-1}$ ($\sim 1.24 \text{meV}$). Subsequent zero-filling of the interferogram for the FFT algorithm resulted in an apparent resolution of $\sim 9.8 \text{cm}^{-1}$ ($\sim 1.22 \text{meV}$) between 0 cm$^{-1}$ and 5000 cm$^{-1}$. The number of averaging scans taken at each temperature also varied on a temperature by temperature basis, with four being the most.

### 3.3.5.2 Results and Discussion

The results from each InSb sample are shown in Figure 3.9 (ME1655) and Figure 3.10 (ME1681) at the end of this section. Although the chosen temperature increments for each sample are not identical, comparing their PT lineshapes at like temperatures good agreement is found. At low temperature the lineshapes appear to show more structure than a typical TDFF lineshape (more clearly seen for ME1655), but no obvious extended FKOs are observed either. It is, therefore, believed these measurements lie in the low- to intermediate-field regime, giving lineshapes with a hybrid character somewhere between TDFF-like and FKO-like (see section 2.3.1). Lineshapes in the intermediate-field regime can often be approximated theoretically by a modified Airy function. However, this approach was not successful in this case. It may be noted though, that at low temperature the lineshapes are approximately symmetric, with the largest amplitude negative lobe located close to its centre. Assuming a strong TDFF-like component to its form, this point can provide a reasonable estimate of the critical point energy, whilst simultaneously providing a common point of reference at each temperature. It is noted that the Varshni behaviour suggested by Vurgaftman and Meyer (also plotted in Figure 3.9 and Figure 3.10 for reference purposes) passes consistently on the long wavelength side of the PT feature, leading to an obvious discrepancy between the estimated critical point energies and accepted bandgap values.[9] This is illustrated in Figure 3.7, where the Varshni fits to each sample’s temperature dependences are also plotted (see appendix, section 9.1 for details regarding the Varshni equation). The resulting fit parameters are listed in Table 3.2.
Table 3.2 Varshni Fit Parameters for both InSb samples studied in this work. Also given are the literature values suggested by Vurgaftman and Meyer.[9]

<table>
<thead>
<tr>
<th>Source</th>
<th>$E_g(T=0)$ (meV)</th>
<th>$\alpha$ (meVK$^{-1}$)</th>
<th>$\beta$ (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME1681 fit</td>
<td>241</td>
<td>0.44</td>
<td>300</td>
</tr>
<tr>
<td>ME1655 fit</td>
<td>243</td>
<td>0.44</td>
<td>326</td>
</tr>
<tr>
<td>Vurgaftman and Meyer [9]</td>
<td>235</td>
<td>0.32</td>
<td>170</td>
</tr>
</tbody>
</table>

Figure 3.7 The temperature dependences of the inferred PT critical point energies for ME1655 (red dots) and ME1681 bulk InSb (blue dots). The corresponding Varshni fits are also shown, along with the Varshni behaviour suggested by Vurgaftman and Meyer.[9] Both samples exhibit similar dependences, with low temperature bandgaps ~6-7meV larger than the accepted value for intrinsic InSb. This difference is believed to be a result of the compressive strain induced in the InSb epilayers due to lattice mismatch with their GaSb substrates. Note the 300K point in the dependence of ME1655 is the result of an absorption measurement (see Figure 3.8).

Note the 300K point for ME1655 is taken from an absorption measurement performed using the fast-scanning FTIR instrument described section 1.1.3.3. That measurement is shown in Figure 3.8. Assuming the absorption coefficient to be proportional to the joint density of states (see appendix, section 9.4), a linear extrapolation of the edge-like feature in a plot of $\alpha^2$ as a function of photon energy indicates the bandgap. The derived energy fits the observed trend in the PT data points (as shown in Figure 3.7)
extremely well, thus showing good consistency with the critical point energies inferred from the PT spectra.

![Graph showing absorption measurement](image)

**Figure 3.8** The room temperature absorption measurement made on sample ME1655. The green curve shows the percentage transmittance, $T$, of the sample. To obtain the bandgap energy, the square of the absorption coefficient, $\alpha^2 = [(\ln(1/T)/x)]^2$ where $x$ is the sample thickness (~2.25μm), as a function of photon energy is plotted (see inset). Because the transmittance does not stabilise at 100% at wavelengths longer than the bandgap, it was necessary to subtract a constant background from $\alpha^2$ (~5.6e7cm⁻²). A linear extrapolation of the edge to zero (red line on the inset) indicates the bandgap energy. A value of $E_g \approx 179.3$meV is obtained, consistent with the PT critical point energies.

The two samples exhibit Varshni dependences with very similar curvatures, separated in energy by ~3meV on average. Judging by the spread of the points in Figure 3.7, the sample temperature dependences are indistinguishable within error margins. At low temperature, the measured bandgap energies of ME1655 and ME1681 are around ~6meV and ~8meV larger than the suggested literature values for intrinsic InSb respectively. As temperature is increased, their dependences appear to converge with that of the literature, reducing that separation to ~2meV and ~6meV respectively at 300K. It may be noted that the PT feature critical point energies become harder to evaluate at higher temperatures due to a rapid weakening of the signal, thus contributing to the increased scatter of points for ME1681 at 250-300K.
Figure 3.9 and Figure 3.10 indicate an evolution of the experimental lineshape from a more FKO-like feature at low temperature to a simpler more TDFF-like feature at high temperature. An evolution of PM lineshapes as a function of temperature has been documented previously, though it is usually noted in those cases that FKOs become weaker at low temperatures.[10,11,12] In doped samples, the field being modulated is due to electrostatic repulsion between electrons and holes, which leads to the formation of a depletion region. At low temperature, electrons and holes become frozen out of the bands, neutralising their donor acceptor states. The depletion region thus becomes larger, which reduces the effective size of the in-built field available for modulation in PM and, hence, $h\Theta$ in equation 2.26 also. However, in this experiment the opposite trend with temperature is observed (i.e. FKOs appear to get stronger at low temperature).

To explain this observed behaviour, one can again refer to equation 2.26 in section 2.3.1, noting that (as explained therein) the various field regimes can be equally well explained in terms of the broadening parameter, $\Gamma$ (i.e. high-field=small $\Gamma$). Thus as temperature is increased, it is changes in $\Gamma$ which dominate the response of $\Gamma/ h\Theta$ in equation 2.26, and the damping of any FKOs is increased. A plausible mechanism for such an increase in $\Gamma$ has been identified in NGSs in the form of a rapidly increasing Auger recombination rate as a function of bandgap energy (and, therefore, temperature). For example, in the case of the CHCC process, the Auger recombination coefficient $C$ is proportional to the Boltzmann factor $\exp(-E_{\text{Aug}}/k_BT)$, where $E_{\text{Aug}}$ is the Auger activation energy equal to $m_eE_g/(m_e+m_h)$ (see section 7.2.1 where Auger recombination is discussed in more detail). Re-labelling the ratio $E_{\text{Aug}}/k_BT=x$, the rate of change of the Auger recombination as a function of $x$ is simply the Boltzmann factor multiplied by -1. Inputting the appropriate bandstructure parameters for InSb and GaAs into the Boltzmann factor, one can obtain a number representing how sensitive the Auger recombination rate is to changes in temperature in those materials.[9] Doing this for $T=300K$ yields a value of 0.819 for InSb and 0.0024 for GaAs, indicating CHCC Auger processes are 346 times more sensitive to changes in bandgap (as induced in the experiment by changes in temperature) in InSb than in GaAs.
This result implies that in NGSs such as InSb the broadening parameter should change far more over a given temperature range than in WGSs. This may explain why a trend of this nature has not been reported in the literature previously, as this is (to the author’s knowledge) the smallest bandgap material to be investigated as a function of temperature using a PM technique to date. This raises the possibility that the observed evolution of the PT signal in these InSb samples is an intrinsic property of such field modulation signals in the regime of NGSs, with the effect appearing more prominent, the smaller the studied material’s bandgap is.

Furthermore, the low mobilities of these particular InSb samples should lead to shorter carrier lifetimes at all temperatures. The more rapid increase in broadening parameter described above should, therefore, occur on top of an already large constant value, equivalent to a small equivalent field (i.e. $\hbar \Theta$) under ambient conditions.

It is, therefore, believed that the strain in the epilayer can provide an explanation for the observed blue shift of the bandgap, while the evolution of the PT lineshapes as a function of temperature is a peculiarity of NGSs; a result of the high sensitivity of Auger recombination rates in such materials to changes in bandgap.
First Application of FTIR Spectroscopy to Photomodulation Techniques

Figure 3.9 PT spectra for ME1655 at a range of temperatures between 77K and 150K. The critical point energies are estimated to coincide with the central negative lobe that dominates the lineshape at low temperature. The literature InSb bandgap values (black line and red dots) consistently lie to the long wavelength side of the PT features. [9] It is believed the apparent blueshift in bandgap is a result of the strain induced by the lattice mismatch between epilayer and substrate.

Figure 3.10 PT spectra for ME1681 at a range of temperatures between 77K and 280K. Signals are observed up to higher temperatures than for ME1655 (Figure 3.9). The observed lineshape evolution is believed to be due to an increase in the broadening parameter, $\Gamma$, as the temperature is increased; a result of a rapid increase in Auger recombination due to the redshift in bandgap. Literature bandgap values lie at longer wavelength than the PT features once again. [9]
3.3.6 Conclusions and Future Work

Using a step-scan FTIR spectrometer, PM signals have been successfully measured in a wide range of material/heterostructures. Where possible the spectra have been compared with those obtained using a pre-existing grating system with generally favourable results (e.g. 1.5-3.5\textmu m). Moving beyond the reach of the grating system, PT has been performed on a pair of similar InSb epilayers (e.g. 5-6.5\textmu m). An estimation of the critical point energies of those PT lineshapes has yielded bandgap energies consistent with expectation. An anomalous lineshape evolution has been observed, however, with the increase in Auger recombination rates in NGSSs cited as a possible explanation. This result indicates that when studying materials lying in this regime, one can expect to observe PM signals exhibiting novel effects not encountered in WGSs. In order to test this hypothesis, it is hoped that more bulk materials in this regime (including more InSb samples) can be studied in the near future.

These measurements indicate a lower wavelength limit for the FTPR system's effective operation at ~1200nm. However the upper limit is in principle bounded only by the transmission of the KBr beamsplitter, which extending extends towards FIR wavelengths (i.e. ~25\textmu m). Of course, the effective use of PM techniques must be taken on a sample-by-sample basis, particularly at MIR wavelengths where many of the samples tested gave no signal whatsoever. Despite this fact, the InSb result in this work shows a hard upper wavelength limit for PM signals (from semiconductors) has not yet been reached. To the author's knowledge, the InSb Pt result also represents a new benchmark for long wavelength temperature dependent PM spectroscopy measurements. That record stood previously at a wavelength corresponding to the room temperature bandgap of InAs (i.e. ~3.5\textmu m), measured in PR mode using a grating spectrometer by Lin et al.[1]

Having provided a hopefully convincing proof of principle, there is a great deal of potential for future work. With the spectrometer's upper wavelength limit not yet reached, one can of course attempt characterisation of longer wavelength materials (or devices). For example, already underway at Surrey is a collaboration with Qinetiq aiming to develop the InSbN alloy system.[13,14] There remain many growth-related
issues in need of addressing before material of sufficient optical quality for reliable PM measurements is produced. However, it is hoped when those difficulties are overcome that the FTPM arrangement can play a key role in any characterisation efforts.

Also, application of the system to lower dimensional structures based on NGSs materials should be possible. In a parallel spintronics study a range of InAs and InSb quantum wells have been grown in order to investigate the effects of quantum confinement on spin lifetimes. In order to fully reconcile any results with theory it would be desirable to experimentally measure the relevant transition energies in the wells. This is a task for which the FTPM arrangement should be well-suited.

Finally, PM is of course not the only modulation technique applicable with the step-scanning FTIR system. With more time it was hoped ER could be attempted, particularly on those samples which gave no PM. Thus if such PM measurements failed simply because the sample in question's in-built fields was too small to generate a significant modulation, ER may yet prove more fruitful. Unfortunately, contacted ER is complicated at MIR wavelengths due to the requirement that the contacts used be transparent to near-bandgap radiation. However, at Surrey the apparatus required for a contactless configuration is available, consisting of a grid-like electrode which sits just above the sample surface. Such a configuration is not ideal for two reasons; firstly the capacitance of the air gap between grid and semiconductor requires the application of very large voltages (~kV) to create a reasonable field in the sample, and secondly the grid itself still blocks a significant fraction of the probe intensity. Nevertheless, it has been successfully tested on the grating PR system, albeit yielding smaller signals than PR in that instance\(^9\), and merits further investigation in conjunction with the FTIR-based arrangement. Ultimately, an upgrade of the spectrometer is desirable, since much of the noise observed in the PM signals is believed to be a symptom of instrumental instabilities. It is firmly believed that by using a (suitable) commercial step-scanning instrument, the SNR in the spectra presented here could be improved by several orders of magnitude.

---

\(^9\) This observation implies that on this occasion the sample possessed a sizable built-in electric field, larger than that generated within the sample by the grid (even when kV were applied). However, for samples with no built-in field a weak signal from ER is obviously preferable to no signal from PR.
3.4 References


4 Bandgap Reduction in Dilute Nitride Materials Based on Narrow Gap Semiconductors

4.1 Introduction

It has been found in III-V semiconductor alloys, that partially replacing the group V element with dilute concentrations of nitrogen has a profound effect on electronic properties. For example, quite extreme reductions in bandgap of the order of 100meV per molar fraction of nitrogen have been observed in GaN$_x$As$_{1-x}$ for $x <$ 0.015. [1] This property of 'dilute nitride' materials in particular has generated a great deal of interest, due to the possibility of using well understood III-V material systems, like GaAs, as a basis for developing long wavelength NIR optoelectronic devices (i.e. in the 1.3-1.5µm wavelength region of interest for telecommunications applications). [2]

In addition to bandgap reduction, the crystal lattice also typically undergoes significant structural relaxation, as nitrogen is generally a considerably smaller atom than the group V element it replaces. As a result, the lattice constant (or unit cell dimension), $a$, is significantly reduced by adding nitrogen. This has implications for lattice matching with standard III-V substrates, as only so much N can be added to the alloy before poor material quality (due to dislocations arising from excessive strain) becomes an issue. In the case of GaNAs, this problem may be solved by adding In to the alloy, which alters $a$ in the opposite direction, whilst still reducing the bandgap from that of GaAs. [3,4] By fine tuning the Ga, In, As and N molar fractions, $x$ and $y$, in Ga$_{1-x}$In$_y$N$_x$As$_{1-x}$, a long wavelength material which remains lattice matched to
GaAs may be obtained (miscibility issues not withstanding).[5] The strongly non-linear effect of N incorporation on bandgap and lattice constant in some common III-V material systems, including GaAs, is illustrated in Figure 4.1. The favoured model used to describe the mechanism responsible for the observed properties of dilute nitrides is the band anti-crossing model (BAC).

![Figure 4.1](image_url) Bandgap bowing in a small selection of III-V-N semiconductor alloys. Adding nitrogen is shown to decrease the lattice constant, \( a \), whilst, in dilute amounts, dramatically reducing the bandgap. Solid lines are the regime of existing experimental data. Dotted lines represent extrapolations (based on best available theory) of this experimental data.[6]

### 4.2 Theory

#### 4.2.1 The Band Anti-Crossing Model

First proposed by Shan et al to explain the pressure and composition dependences of the PR spectra of GaInNAs, this particularly elegant model has been enormously successful in reproducing the observed conduction band structure of several wide gap dilute nitride materials (e.g. GaAsN, GaInNAs, GaPN, InGaN etc.).[7,8] As indicated in Figure 4.1, efforts have already been made to extend its applicability to nominally NGSs such as GaSb, InAs, and InSb in the hopes of producing a class of optoelectronic materials capable of covering the entire MIR wavelength range (i.e. 2-15\( \mu \)m).[9-13] However, optical and electrical characterisation of these materials has proven a formidable challenge, to say nothing of the challenges encountered in their
growth. For this reason verification of the model in the MIR regime remains elusive.[14]

In trace quantities, nitrogen contributes acceptor-like levels to the bandstructure of III-V materials. Within the framework of the BAC model, dilute nitrogen concentrations give rise to a narrow band, reflecting the highly localised nature of the sparsely populating and highly electronegative nitrogen atoms\textsuperscript{10}, found close in energy to the conduction band minimum. Due to the proximity of the localised nitrogen band to the extended conduction band states, there is a weak mixing interaction between them. In accordance with $k.p$ theory, the energies of the resulting mixed states may be obtained by solving the following eigenvalue problem:[7,8]

$$\begin{vmatrix}
E - E_M(k) & -V_{NM} \\
-V_{NM} & E - E_N
\end{vmatrix} = 0$$

(4.1)

where $E_M(k)$ are the energies of the extended conduction band states, $E_N$ the energy of the localised nitrogen band, and $V_{NM}$ describes the coupling interaction between the two. Computing this determinant and rearranging to obtain a quadratic function of $E$, the quadratic formula reads:

$$E_{\pm}(k) = \frac{(E_N + E_M(k)) \pm \left[(E_N - E_M(k))^2 + 4V_{NM}^2\right]^{1/2}}{2}$$

(4.2)

The coupling interaction thus splits the conduction band into two sub-bands, $E_-$ and $E_+$, separated in energy by $\Delta E = 1/2[(E_N - E_M(k))^2 + 4(V_{NM})^2]^{1/2}$. When the nitrogen states are resonant with the minimum of the conduction band (i.e. $E_N = E_M(k=0)$), the separation at the $\Gamma$-point reduces to $\Delta E = V_{NM}$, with the $E_-$ and $E_+$ split equally by energy $V_{NM}/2$ about $E_N$. The square of the matrix element coupling the extended and localised states is given by:[7,8]

\textsuperscript{10} A narrow band corresponds to an extremely large effective mass. Electrons with large effective masses are correspondingly difficult to accelerate or move ($v$ is proportional to $dE/dk$). Hence, they tend to remain trapped (or localised) around the nitrogen sites.
where $|k\rangle$ has an extended Bloch-like form and $|N\rangle$ has a localised wavefunction form. Equation 4.3 is only valid for low nitrogen concentrations as interactions between neighbouring nitrogen atom have been neglected. If one also assumes the substitutional nitrogen atoms are randomly distributed through the host lattice it may reasonably be expected that $|V_{NM}|^2$ scales in proportion with the molar fraction of nitrogen in the alloy. Indeed, it can be shown that $|V_{NM}|^2$ may be expressed as the product of a material dependent constant term, $|C_{NM}|^2$, and x, yielding:

$$|V_{NM}|^2 = \langle k|V|N\rangle \langle N|V|k\rangle$$  \hspace{1cm} (4.3)

Therefore, the interaction between $|k\rangle$ and $|N\rangle$ generally results in two highly non-parabolic sub-bands whose characters are a mixture of $N$ and $k$. Which fraction of each goes into each sub-band depends on the proximity of $E_N$ and $E_M(k=0)$. The $E$ sub-band defines a new conduction band, the minimum of which is found at lower energy than $E_M(k=0)$, thus explaining the redshift in bandgap observed in experiment. In addition to the two transitions between this conduction band minimum and valence band/split-off band, the BAC model also predicts two new equivalent transitions associated with $E_+$. It was observation of these transitions, along with their predicted pressure and composition dependences, in the PR of Shan et al that initially provided the most convincing evidence of the BAC models validity.[7,8]

$$V_{NM} = C_{NM}x^{1/2}$$  \hspace{1cm} (4.4)

Figure 4.2 shows the bandstructure of InAs$_{1-x}$N$_x$ for $x=0.02$, as predicted by equation 4.2. Model parameters are based on calculations by O’Reilly and Lindsay where $V_{NM}=(-1.3)x^{1/2}$, $E_N=1.36eV,[12]$ while the values of $E_M(k)$ are based on Vurgaftman and Meyer’s suggested InAs bandstructure parameters.[19] In Figure 4.2 it is clearly seen that unlike GaInAsN and GaPN, where $N$ state energies are close to resonance with the conduction band minimum, here $E_N$ lies some way above $E_M(k=0)$. In fact it is $\sim 1eV$ above, almost three times the size of the InAs bandgap. As a result, close to the $\Gamma$-point the $E$ sub-band retains an approximately parabolic shape (small effective mass enhancement), while the $E_+$ is distorted considerably from the parabolic (large effective mass enhancement). In other words, the character of $E_+$ is close to that of the
InAs conduction band, or $|k\rangle$ (neglecting the intrinsic non-parabolicity of the InAs conduction band), whereas $E_\pm$ has strong localised, or $|N\rangle$, character. Because oscillator strengths between extended states are greater than those between localised states, at the $\Gamma$-point transitions associated with $E_\pm$ are expected to dominate.

![Band structure diagram](image)

**Figure 4.2** Basic BAC calculation of InAs$_{0.98}$N$_{0.02}$ bandstructure. The induced non-parabolicity of the $E_\pm$ subbands is clear (the intrinsic non-parabolicity of the InAs conduction band has been neglected). Also of note is the large gap between $E_N$ and $E_M(k=0)$, which results in a smaller value of $V_{NM}$ at the band edge per molar fraction of nitrogen than for GaAs (for example).

The BAC model as presented thus far assumes that $E_N$ and $E_M(k=0)$ are unaffected by the molar fraction of nitrogen incorporated. At low nitrogen concentrations this is reasonable, to a first approximation. However, from a purely statistical standpoint, eventually the nitrogen concentration can become large enough that a significant proportion of the nitrogen-related sites in the material are complexes of more than one neighbouring nitrogen atom. Obviously the first kind of complex to form in this way is a nitrogen pair, the self-energy of which is generally lower than $E_N$ (the self-energy of isolated nitrogen atoms in the preceding discussion). Therefore, as nitrogen is added to an alloy $E_N$ does not remain constant, exhibiting rather a gradual evolution between isolated to pair-like. If one assumes this evolution to be linear with respect to $x$ it can be accounted for by introducing an additional parameter, $\gamma$, into the model, giving the following form for $E_N$:[15]
$E_N = E_N^0 - \gamma x$  \hspace{1cm} (4.5)

Where $E_N^0$ is the isolated nitrogen level energy (formerly $E_N$). This expression’s validity is contingent on minimal interaction between neighbouring complexes (i.e. no overlap of their wavefunctions), and so applies only as far as the regime of pair formation. By the time nitrogen concentrations conducive to the formation of higher order complexes\textsuperscript{11} are reached, this condition is violated, and equation 4.5 is rendered invalid. Tight-binding calculations by O’Reilly et al on GaAsN related-alloys indicate this breakdown occurs for $x \sim 0.05$, well above the highest sample concentration in this work (i.e. $x < 0.025$).\textsuperscript{[15]}

Besides the mixing interaction that is the BAC model, the incorporation of nitrogen has a secondary (though smaller) effect on $E_M(k)$. This can be thought of as a transition between the ‘ideal’ crystalline host material to one where the crystal periodicity is interrupted up by randomly occurring nitrogen impurities. The conduction band minimum thus tends to reduce with increasing nitrogen composition, an effect often referred to as bandgap shrinkage (or narrowing). Again assuming the variation to be linear in $x$, the conduction band can be described as follows: \textsuperscript{[15]}

$$E_M(k) = E_M^0(k) - \alpha x$$  \hspace{1cm} (4.6)

where $E_M^0(k)$ is the intrinsic host alloy conduction band (formerly $E_M(k)$), and $\alpha$ is the parameter describing the bandgap shrinkage. Including these additional effects equation 4.1 may be re-written as follows:

$$\begin{bmatrix}
E - E_M^0(k) - \alpha x & -\beta x^{1/2} \\
-\beta x^{1/2} & E - E_N^0 - \gamma x
\end{bmatrix} = 0$$  \hspace{1cm} (4.7)

where the coupling interaction is here described by the parameter $\beta$, which is equivalent to $C_{NM}$.

\textsuperscript{11} Complexes of up to four nitrogen atoms bonded to a single In atom can be formed, corresponding to substitution of every nearest neighbour As atom.
The bandstructure predicted by equation 4.7 for InAs$_{1-x}$N$_x$ is illustrated in Figure 4.3 for a range of compositions; $x=0.005$, $x=0.02$, $x=0.04$, and $x=0.08$. The model parameters are the same as those used in Figure 4.2, only with additional variables $\alpha=1.3\text{eV}$ and $\gamma=2.0\text{eV}$ ($\beta x^{1/2}=V_{NM}$). The BAC model is thus shown to predict a redshift in the InAsN bandgap (relative to InAs) of $\Delta E_g=59\text{meV}$ for a moderate nitrogen composition of $x=0.02$; a $\sim17\%$ reduction in the bandgap. This corresponds to a shift in wavelength from $3.51\mu\text{m}$ to $4.28\mu\text{m}$. Although the large difference in $E_N$ and $E_M(k=0)$ means this shift is smaller than those in wider gap host alloys for similar nitrogen compositions, it is still considerable.

![Figure 4.3 BAC prediction of the variation of the $E_-$ and $E_+$ sub-bands in InAs$_{1-x}$N$_x$ as function of composition, including the effects of bandgap shrinkage and nitrogen complex formation. Note that as composition is increased from $x=0.005$ to $x=0.08$, the bandgap (defined by $E_-$) is reduced in a non-linear manner (as indicated by the orange-dashed). The redshifts in bandgap are $\Delta E_g \sim 15\text{meV}$ ($x=0.005$), $\Delta E_g = 59\text{meV}$ ($x=0.02$), $\Delta E_g = 117\text{meV}$ ($x=0.04$), and $\Delta E_g = 230\text{meV}$ ($x=0.08$).]

Also visible in Figure 4.3 is the aforementioned enhancement of the effective mass of the $E_-$ sub-band as nitrogen is added. This effect is more clearly illustrated in Figure 4.4, where the effective mass ratio, $m/m_M$, is plotted for each concentration as a function of energy. An analytical expression for that ratio can be obtained using the following well known definition of effective mass:

$$m = \hbar^2 \left( \frac{d^2 E}{dk^2} \right)^{-1} \quad (4.8)$$

Substituting $m=m_-$ and $E=E_-$ (where $E_-$ is defined by equation 4.2) into this expression and performing the differentiation yields:[8]
\[ m(E) = m_M \left\{ 1 + \left( \frac{V_{NM}}{(E_N - E)} \right)^2 \right\} \] (4.9)

where \( E < E_N \). A trivial rearrangement gives the expression for \( m/M \) plotted in Figure 4.4.

![Figure 4.4 The effective mass ratios of \( E \) in \( \text{InAs}_{1-x}\text{N}_x \) as a function of energy for each composition illustrated in Figure 4.3. For small \( x \), the enhancement of \( m \) is significant only at energies close to \( E_N \) (dashed black line). Increasing \( x \) to a few percent enhances \( m \) at energies closer to the conduction band minimum. For \( x=0.02 \) (a moderate composition) this enhancement is \( \sim 30\% \) around \( \sim 300\text{meV} \) above \( E_{MO}(k=0) \). The vertical dashed lines indicate \( E_N \) at each composition (colour-coded to match the relevant curve).

For low compositions it is seen in Figure 4.4 that the enhancement of the \( E \) sub-band’s effective mass is only significant at energies close to \( E_N \), corresponding to very large \( k \). However, at larger compositions (i.e. \( x \approx 0.05 \)), the enhancement becomes increasingly significant at energies closer to the conduction band minimum. This effect is amplified the closer \( E_N^0 \) is to \( E_{MO}(k=0) \) in the material system (i.e. a larger \( V_{NM} \) close to the band edge). As seen in Figure 4.3, the effective masses of the \( E \) and \( E+ \) sub-bands follow opposite trends with increasing composition. Therefore, as nitrogen is added their masses draw closer to equality, until at some composition they
should eventually cross over. Past that point the dominance in optical transitions is expected to shift to those associated with \( E_+ \).[7]

A powerful extension to the BAC model has been provided by O’Reilly et al. By performing tight-binding calculations on randomly generated GaAsN supercells, the interaction between each and every nitrogen atom is treated explicitly using a linear combination of isolated nitrogen states (LCINS).[16,17,18] In so doing, those authors greatly enhanced the predictive capacity of the BAC model, identifying an array of localised nitrogen cluster states which hybridise with conduction band states to produce bandgap reduction and effective mass enhancement. In order to fully quantitatively describe the conduction band dispersion observed experimentally in GaAsN it is necessary to take all these cluster states into account, particularly those close to the conduction band edge. In this respect the LCINS approach is a step beyond the two-level BAC model.

4.2.2 GaSb\(_{1-x}\)N\(_x\)

The bandgap of the host GaSb system lies at the very edge of the MIR regime, with \( E_g(T=0) \approx 812 \text{eV} (1.53 \mu\text{m}) \).[19] For this reason, GaSb is often described as an intermediate gap semiconductor, occupying the wavelength range between visible/NIR materials (e.g. GaAs, InP), and those in the MIR (e.g. InAs, InSb). The room temperature bandgap has also been quite well documented using techniques such as PL, PR, and absorption, the averages of several such measurements placing its value at \( \approx 0.726 \text{eV} (1.71 \mu\text{m}) \).[19] Over recent years GaSb has become an increasingly important component of mid-infrared optoelectronic devices. A thorough review of the materials system’s properties have been provided by Dutta et al.[20]

The electronegativity mismatch between antimony and nitrogen is larger than that of any other combination of commonly used group V elements, leading some to conclude that the nitrogen level in dilute antimonide nitrides might be particularly close to resonance with the conduction band edge.[10,20,21] If so, the bandgap reduction can be expected to be even more extreme in GaSb\(_{1-x}\)N\(_x\) (and other dilute antimonide nitride alloys) than that observed for GaAs\(_{1-x}\)N\(_x\) (and other dilute arsenide nitrides). Lattice matching of the nitrogen containing alloys with GaSb is possible
Bandgap Reduction in Dilute Nitride Materials Based on Narrow Gap Semiconductors

(along with further reduction in the bandgap) by adding In, producing Ga_{1-x}In_xSb_{1-x}N_x. The predicted variation of bandgap and lattice constant with nitrogen composition for some common antimonide nitride materials is shown in Figure 4.5.

![Figure 4.5 Variation of bandgap and lattice constant as a function of nitrogen composition for dilute antimonide nitride alloys, as compared to GaAsN. The bowing effect is more extreme in GaSbN than GaAsN. Note also that the bandgap of InSbN reduces to zero with just a few percent of added nitrogen by virtue of its already very narrow gap.[6]](image)

4.2.3 InAs_{1-x}N_x

Host alloy InAs has been a key material for MIR optoelectronic applications for several decades. As a result, both its electrical and optical properties have been quite exhaustively studied (though not to the extent of GaAs and Si of course).[19] Its room temperature bandgap of \(~0.35\text{eV} (3.51\text{\textmu m})\) lies conveniently within an atmospheric absorption window, and provides a stepping stone between GaSb and InSb (see Figure 4.1), with which it forms the material backbone of MIR applications. However, the proclivity of InAs to \(n\)-type conductivity has made unambiguous confirmation of bandgap reduction in InAsN using optical absorption techniques troublesome.[22,23] This is because such unintentional degenerate \(n\)-type doping results in a Moss-Burstein blue-shift of the optical absorption edge of the material. Even if this shift is small in the host InAs, reducing the bandgap in InAsN alloys serves to exacerbate any problem. In order to extract the true bandgap reduction from absorption measurements, the contribution of the Moss-Burstein shift must be modelled and subtracted. With significant uncertainties surrounding the applicability of the BAC
model to NGS-based dilute nitrides, in doing so one must use unproven assumptions about the material bandstructure to determine the material bandstructure; an argument which is clearly circular and prone to error. This approach was taken by Shih et al, however, with their absorption measurements initially indicating a blue shift in bandgap before correction.[12] Blue shifts with increasing nitrogen composition were also found by Kuroda et al, though extraction of the true bandgap was in these cases beyond the scope of the work presented.[13] Only one result (Naoi et al) has claimed unambiguous observation of a reduction in InAsN bandgap from absorption measurements.[24]

What is ideally required is a technique which can observe any redshift in bandgap directly, irrespective of Moss-Bursetin shift. This is a task for which PL should be well suited, as radiative recombination predominantly occurs between carriers at the band edges.[25] Specifically, assuming the Maxwell-Boltzamann approximation holds for carrier occupations, within $k_B T$ of the band edge. It may be noted that for large carrier concentrations band-filling effects modify the PL emission lineshape, as described in more detail in section 2.2. However, so long as fitting of the emission peak uses the appropriate form for the occupation function, the actual bandgap energy may be extracted from an emission peak. Of course, such measurements require samples of good optical quality to prove successful. Unfortunately, the high carrier concentrations expected in InAsN alloys are expected to increase Auger recombination rates (see section 7.2.1), reducing the materials radiative efficiency beyond the already relatively low values encountered in InAs. A primary aim of these measurements was, therefore, to investigate the viability of PL as a means of directly observing bandgap reduction in InAsN. If a shift was observed, the intent was then to investigate that reduction over a range of sample compositions. As discussed in section 4.2, the BAC model predicts any such reduction to be less extreme than that observed in GaAsN (or predicted for GaSbN), though it is still far from inconsequential. The bandstructure of InAsN for $0.005 \leq x \leq 0.08$ is illustrated in Figure 4.3.

The fact that PM signals arise from modulation of a sample’s absorption edge(s) means such techniques are unlikely to be of any help in observing band edge
transitions in InAsN; the same Moss-Burstein shift-related limitations apply to PM as do absorption measurements.

**4.3 Experiment I: GaSb$_{1-x}$N$_x$**

**4.3.1 Samples**

The GaSb$_{1-x}$N$_x$ samples used in this work were grown by Qinetiq using nitrogen plasma-assisted molecular-beam epitaxy (MBE), the details of which may be found elsewhere.[44] A variety of alloy compositions were produced, including Ga$_{1-y}$In$_y$Sb$_{1-x}$N$_x$, with nitrogen incorporations as high as $x=0.015$ believed to have been achieved based on x-ray diffraction (XRD) measurements. Fractions of In as high as $y=0.6$ were added to alloy compositions with $x=0.015$. However, at such low compositions an approximately 1:4 ratio of $x:y$ should maintain lattice matching, so one might expect the best quality samples when $y=0.06$. That is close to the smallest non-zero In composition grown; $y=0.07$. Table 4.1 gives a summary of the key samples used in this study. Henceforth, specific samples shall be referred to by their sample code as listed in Table 4.1.

Effectively two groups of GaSbN sample were grown; one with thin film epilayers (~150nm) on (approximately) lattice matched GaSb substrates, and one with thicker epilayers (~2µm) on mismatched GaAs substrates. There were several reasons for taking this approach. Firstly, it was of primary importance to determine the quality of the epilayer crystal growth using XRD, an assessment best made on unstrained layers (i.e. GaSb substrates). However, with the ultimate intention of the work being to perform optical characterisation of the material (without precluding the possibility of transmission/absorption measurements), a substrate with a bandgap well above that of the epilayer was desirable. The obvious candidate for this was GaAs ($E_g\sim1.43$eV, or 0.87µm). The lattice mismatch between GaAs and GaSb induces considerable strain in the epilayers of such samples though, possibly introducing a number of undesirable strain-related problems (e.g. blueshifts in $E_g$, dislocations, etc.). By comparing XRD from similar epilayer compositions grown on each substrate (i.e. assuming the epilayers on GaAs are 100% relaxed), the effect of strain on those samples grown on GaAs could thus be evaluated. Such measurements were performed by fellow
researchers at the University of Warwick, and indicated epilayers grown on GaAs were ~97-98% relaxed.

<table>
<thead>
<tr>
<th>Sample Code</th>
<th>Structure (epi/sub)</th>
<th>Epilayer Thickness (μm)</th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate</td>
<td>GaSb (ud)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RF46</td>
<td>GaSb (ud)/GaAs</td>
<td>1.50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RF25</td>
<td>GaSb₁₋ₓNₓ/GaSb</td>
<td>0.15</td>
<td>0.005</td>
<td>-</td>
</tr>
<tr>
<td>RF30</td>
<td>GaSb₁₋ₓNₓ/GaSb</td>
<td>0.15</td>
<td>0.010</td>
<td>-</td>
</tr>
<tr>
<td>RF31</td>
<td>GaSb₁₋ₓNₓ/GaSb</td>
<td>0.15</td>
<td>0.015</td>
<td>-</td>
</tr>
<tr>
<td>RF44</td>
<td>GaSb₁₋ₓNₓ/GaAs</td>
<td>2.00</td>
<td>0.005</td>
<td>-</td>
</tr>
<tr>
<td>RF43</td>
<td>GaSb₁₋ₓNₓ/GaAs</td>
<td>2.00</td>
<td>0.010</td>
<td>-</td>
</tr>
<tr>
<td>RF42</td>
<td>GaSb₁₋ₓNₓ/GaAs</td>
<td>2.00</td>
<td>0.015</td>
<td>-</td>
</tr>
<tr>
<td>RF47</td>
<td>Ga₁₋ₓInₓSb₁₋ₓNₓ/GaAs</td>
<td>2.00</td>
<td>0.015</td>
<td>0.070</td>
</tr>
</tbody>
</table>

Table 4.1 The samples investigated in this study. Substrate and RF49 are control samples. The epilayers of samples RF42, RF43, and RF44 (grown on GaAs substrates) were determined from XRD measurements to be 97.1%, 97.9%, and 97.6% relaxed respectively.

### 4.3.2 Experimental Procedure

PR and PL have been performed on the samples listed in Table 4.1 using the Czerny-Turner grating spectrometer described in section 2.2. The rest of the set up was essentially the same as the typical arrangements detailed in section 2.4. The samples of size ~5×5mm were mounted on copper plates using silver conductive paint, and placed on the cold finger of a Cryophysics closed-cycle He cryostat capable of reaching a minimum temperature of ~8K. The only available lid for this cryostat had a cylindrical glass window, leading to some absorption at MIR wavelengths. Though not a major issue at wavelengths corresponding to the GaSb bandgap, for x=0.015 a redshift to ~0.36eV (3.5μm) was anticipated, close to a prominent absorption band in the window at ~2.8μm. However, the only other available cryostat was of nitrogen bath design. Although the windows of that model could be changed for windows more suitable for MIR wavelengths (e.g CaF₂), it was decided that the ability to go below 77K with the He cryostat was more useful on this occasion.

For the PL measurements a diode pumped Nd:YAG laser was used as the excitation
Bandgap Reduction in Dilute Nitride Materials Based on Narrow Gap Semiconductors

source, with maximum power output ~100mW at a wavelength of 1064nm. The spot size on the sample was ~1mm in diameter, giving a power density of approximately 10Wcm\(^{-2}\) (actually quite low compared to other similar studies).[26-29] Modulation was performed at audio frequencies by a mechanical chopper, the detector output then being processed by a lock-in amplifier in an effort to improve the SNR. A 1200nm long pass filter was used to remove laser scatter when measuring bulk GaSb. Unfortunately no optical filters suitable for wavelengths in the 3-6\(\mu\)m region were available. Instead a piece of GaSb wafer with known optical properties was used to block laser scatter for GaSbN samples whose bandgaps are predicted to lie in that range (typically spectra were collected with and without the GaSb wafer to make sure nothing was being missed).

PR measurements were performed using a ~10mW HeNe laser as the modulation source. Modulation frequencies ~800 Hz were used. At frequencies lower than that, vibrations induced by the compressor/pump combination of the cryostat led to noise signals too large for the lock-in amplifier to effectively filter out. Above that frequency a gradual reduction in \(\Delta R\) signal was observed, decreasing by around one-half between 875Hz and 1600Hz (the chopping frequency dependence of PM signals was discussed in section 2.3.2).

The same liquid nitrogen-cooled InSb detector was used for both PL and PR measurements. This detector’s matched pre-amp was ac-coupled, so in order to obtain the dc component of the signal for \(\Delta R/R\), a separate Stanford filter/amplifier with variable filter bandwidths, including the option of no filtering, was employed.

4.3.3 Results and Discussion

The results are divided into two parts; those obtained on the binary GaSb material (PL and PR), thereby providing validation of the experimental procedure and an experimental control, and any results obtained from the GaSbN epilayers. Each of these parts is further subdivided into descriptions of any PL and PR results.

All quoted/plotted PL transition energies may be assumed to be the peak energy less \(k_BT/2\), unless otherwise stated (see section 2.2.1). Fitting of PR spectra (where
relevant) was performed using the ‘marqu’ lineshape fitting program designed and written by Dr. T. J. C. Hosea.

4.3.3.1 Bulk GaSb

4.3.3.1.1 Photoluminescence

RF46
The PL from RF46 as a function of temperature is shown in Figure 4.6. Signals were only measurable up to 150K. The accepted literature Varshni behaviour (see appendix, section 9.1) is also shown in Figure 4.6.[19] Notably, the observed emission occurs well away from the band edge at all the sampled temperatures. An emission band at this energy has been observed by others previously, where it was attributed to free-to-bound transitions involving a deep acceptor level (labelled T in that work).[26,27] Though the origin of the feature was not solidly identified, the authors suggested a complex defect consisting of a Te donor and a stoichiometric defect might be responsible. As sample RF46 is not intentionally doped there should be no Te donors present, casting some doubt on either that assignment or this sample’s growth.

Substrate
As seen in Figure 4.7, the PL emission from the GaSb substrate material is somewhat better behaved that that from RF46, with measurable signals obtained at 300K. Although at 15K the emission still appears considerably below the band edge, it is in broad agreement with existing results.[26] In that work, the most prominent transition (labelled A) in the low temperature PL of nominally undoped material was attributed to recombination at a singly-ionised native double acceptor. There is a slight discrepancy in the energy of this feature, however. Existing literature values include ~774meV (at 2K) [26,27] and ~778meV [28,29] (at 16K), compared to this study’s result of ~782meV (at 15K). The band on the high energy side of A (labelled D) is attributed to recombination of a free hole and an electron bound to a shallow donor.[27] The evaluated transition energy for D of ~798meV is a little lower than the generally accepted values of 805meV [26] and 801meV [28] (at 2K), and 802.9meV at 16K [28,29]. The weak band on the low energy side of A is a commonly observed
complex which includes the LO-phonon replica of A [28,29]. In the 50K spectrum, the onset of A's ionisation can be seen (i.e. the peak becomes weaker as the temperature is increased), while a prominent shoulder develops at ~805meV peaked close in energy to the suggested literature bandgap at each temperature.[19] By 100K the intensity of that feature exceeds that of A. At room temperature, the PL emission (which is then dominated by the feature) is peaked at ~727meV, very close to the accepted room temperature bandgap of ~726meV, and in good agreement with the room temperature PL of Dutta et al.[45] However, the PL transition energy (peak value less \( k_B T / 2 \)) is ~714meV\(^{12} \), well below that accepted value. Therefore, the feature is not believed to be associated with a band-to-band transition. The stoichiometric defect-related feature observed in the PL of RF46 is not present in this spectrum. This in itself should be an indication of good crystal quality.

\(^{12}\) In the substrate sample the Boltzmann approximation should hold. This has been verified by fitting the 300K emission using equations 2.10 and 2.11, which yielded a transition energy approximately equal to the peak energy less \( k_B T / 2 \) (or ~714meV).
Figure 4.6 PL emission from RF46 between 15K and 150K. The observed emission consistently lies at lower energy than the accepted bandgap.[19] The feature has been observed previously by others (labelled T), and was tentatively attributed to a stoichiometric-related defect.[26,27] Grating slit settings (horizontal×vertical dimensions) and intensity scaling factors are indicated.

Figure 4.7 PL emission from GaSb substrate material between 15K and 300K. The observed emission is in good agreement with other sources.[26,45] Grating slit settings (horizontal×vertical dimensions) and intensity scaling factors are indicated. Annotations to peaks indicate band labels in accordance with literature.
4.3.3.1.2 Photomodulated Reflectance

RF46

In stark contrast to the PL emission from RF46, PR from this sample shows a strong TDFF-like PR lineshape, localised very close to the literature bandgap energy.[19] Fitting the temperature dependent data shown in Figure 4.8 with a TDFF lineshape, the transition energies are found to be slightly below the accepted literature dependence (by ~2meV).[19]. However, the general trend of this function is reproduced quite precisely by the experimental results. This can be seen more clearly in Figure 4.9, a plot of the feature’s critical point energy (derived from the TDFF fit) as a function of temperature. The parameters resulting from a Varshni fit to that data are given in Table 4.2.

<table>
<thead>
<tr>
<th>Source</th>
<th>$E_g(T=0)$ (meV)</th>
<th>$\alpha$ (meVK$^{-1}$)</th>
<th>$\beta$ (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF46 Fit</td>
<td>809</td>
<td>0.45</td>
<td>169</td>
</tr>
<tr>
<td>Vurgaftman and Meyer</td>
<td>812</td>
<td>0.42</td>
<td>140</td>
</tr>
</tbody>
</table>

Table 4.2 Varshni Fit Parameters obtained from a fit to the RF46 PR data. Also given are the literature values suggested by Vurgaftman and Meyer for comparison.[19]

The only point of note beyond the quality of this result is the lineshape’s behaviour below 50K. Below that temperature a quite dramatic change in lineshape is observed, with what appears to be a second feature appearing just below the main feature. Any low temperature onset of Franz-Keldysh Oscillations (FKO’s) can be ruled out as a cause, as the secondary oscillation appears at a lower energy than the main feature. Also the energy of this feature is not low enough for it to be related to the acceptor transition (A) which dominates the low temperature PL of the GaSb substrate sample. Rather, the transition energy at 15K appears to match approximately the position of the feature labelled D. Its rapid quenching with temperature is likely due to ionisation of the shallow donor to which the recombining hole is bound. Its anomalous temperature dependence is probably a consequence of errors arising from the fitting of an extremely weak TDFF on top of a much larger one (i.e. the main feature), so the derived energies for the feature may not be trustworthy.
Also plotted in Figure 4.9 is a literature curve based on Varshni parameters from the PR study of Munoz et al.[30] Comparison of that curve with this study’s result shows remarkable agreement at low temperature. At high temperature there is some deviation of the two data sets, however, and those authors do not observe any secondary transition below 50K in their data. In summary, the accepted variation of $E_g$ is broadly reproduced by this work’s PR data, though at low temperature the absolute energies compare slightly more favourably with the PR results of Munoz et al than the PL-based values quoted by Vurgaftman and Meyer.[19] It may be noted that the spectral region containing the stoichiometric defect-related feature in this sample’s PL was investigated using PR, though no feature was found.

**Substrate**

No PR could be observed from the GaSb substrate sample. The apparent inconsistency in the PL and PR of each sample is believed to be associated with the samples’ structures. RF46 is an epilayer sample grown on GaAs, which leads to a large lattice mismatch (as discussed in section 4.3.1). The defect-related states that result, particularly close to the interface region, can trap carriers. In PL, carrier trapping is detrimental because it increases the probability of recombination by Shockley-Reed-Hall processes, reducing the radiative efficiency. Additionally, however, the lattice mismatch at the substrate/epilayer interface may lead to considerable band bending due to their different Fermi-levels. That in-built electric field is most likely the one modulated by the pump laser during PR. In the substrate sample no such interface exists, and so no significant electric field (assuming the surface fields are too weak to produce a modulation signal) is available for modulation. Hence, the GaSb epilayer grown on GaAs may produce strong PR signals (by virtue of the large in-built field at the layer interface), but little or no PL (i.e. due to carrier trapping). Conversely, the substrate may yield little PR (due to the lack of a band bending at the interface), yet produce strong PL (due to its good crystallographic quality). This hypothesis is thus consistent with the measured signals.
Figure 4.8 PR Signal from RF46 between 15K and 300K. The main feature lies very close to the suggested literature bandgap energy at all temperatures. Although these spectra do not extend to the energies which contained the Te defect-related feature in this sample's PL, that spectral region was investigated and no feature was observed. Dashed curves indicate TDFF lineshape fits. Broad background oscillations at 300K prevented fitting of that spectrum. Use of ND filters and intensity scaling factors are indicated, as are transition labels at 15K (same as PL spectra notation).
4.3.3.2 GaSb$_{1-x}$N$_x$ Epilayers

4.3.3.2.1 Photoluminescence

For this part of the study, PL was only attempted on the high nitrogen composition sample of a particular sample group if the lowest composition sample of that group yielded a signal. The groups in question were: thin layers on GaSb, thick layers on GaAs (and thick Ga$_{1-y}$In$_y$Sb$_{1-x}$N$_x$ layer on GaAs). This approach was necessary due to time constraints imposed on the use of apparatus. Unfortunately, using this methodology only three of the samples provided (the thin epilayers layers grown on GaSb substrates) exhibited any PL emission at all, and none of those showed any redshift in their emission relative to GaSb.

RF25

The PL from sample RF25 is shown in Figure 4.10. Also shown there is the PL obtained when this sample was reversed (i.e. PL measured from the substrate). Aside from the relative weakness of the epilayer signal, both epilayer and substrate are seen to produce almost identical PL (note the signal from the substrate has been normalised.
to the same amplitude as that from the epilayer in Figure 4.10. It may also be noted that the feature T seen in the PL of RF46 (at ~700meV) is present in the emission from the epilayer side, but not from the substrate side. Therefore, it is believed that the features A and D originate from the substrate in both cases, but the feature T originates only from the epilayer. This is consistent with emission observed from RF46 (epilayer) and the substrate sample earlier. The relative weakness of A and D in emission from the epilayer side is thus attributed to absorption within the epilayer itself.

**RF31**

The PL from RF31 is shown in Figure 4.11. Clearly there is little or no observed redshift in energy of any of the observed features, as compared to their binary equivalents. The spectrum was investigated at ~3.5μm (at room temperature), where the shifted bandgap was expected to lie based on BAC model calculations, but to no avail. Only weak second order replicas of the higher energy features were found there. Importantly, the feature T is here found to have gained intensity compared to RF25, possibly indicating more heavily defected material. The feature T may, therefore, originate from a purely stoichiometric defect. Its strength in a PL spectrum could thus provide a measure of a sample’s crystal quality (e.g. in this case, as nitrogen is added the crystal is negatively impacted).
Figure 4.10 PL emission from RF25 between 15K and 300K. No redshift in the emission is discernable. The emission from the substrate (i.e. sample reversed) is shown for selected temperatures as a grey dashed line (normalised to the epilayer emission peak). Features A and D are consistent with substrate sample emission, while the feature T is that observed in the PL of RF46.
Figure 4.11 PL emission from RF31 at several temperatures between 15K and 300K. No redshift in the emission is discernable. Features A and D are consistent with substrate sample emission, while the feature T is that observed in the PL of RF46. The feature T also appears stronger here than in the PL of RF25 (see Figure 4.10) possibly indicating an increase in the number of stoichiometric defects as a result of nitrogen incorporation.
4.3.3.2 Photomodulated Reflectance

Unfortunately no PR signal could be obtained from any of the GaSbN epilayers provided. This includes the thicker GaSbN epilayers grown on GaAs substrates; the dilute nitride samples with the most in common with RF46, from which strong PR signals were obtained.

4.3.4 Conclusions and Future Work

A selection of bulk (control) GaSb and GaSbN samples have been studied. PL and PR results on the binary material indicate that the layer thickness and/or substrate material play an important role in these material’s optical properties, with PL observed from substrate material though no PR, and vice versa for an epilayer grown on a lattice mismatched substrate. However, in examining similarly structured dilute nitride samples, the PL observed was found consistent with lower quality examples of the binary material. No energy shifts or unexpected spectral features (i.e. features not seen in PL from GaSb in this or other works) were observed in what PR or PL was measurable from the studied samples.

Unfortunately, the only conclusion that can be derived from these measurements is a confirmation that a considerable amount of work remains to be done to improve the growth of GaSbN material. Presently, the sample quality is too poor to yield optical properties sufficient for optical characterisation. However, despite this negative conclusion, it may be noted that Prof. McConville’s group at Warwick University have obtained what they believe to be significant (though extremely weak) absorption features (associated with $E_e$ and $E_v$) from a batch of samples closely related to the ones studied in this work. They are currently working on theoretical models able to better describe the extreme bandstructure of GaSbN in order to confirm their hypothesis by prediction.
4.4 Experiment II: InAs$_{1-x}$N$_x$

4.4.1 Samples

The InAs$_{1-x}$N$_x$ samples used in this work were grown at Sheffield University by MBE. Details of similar growths may be found elsewhere.[31] Samples of varying epilayer thickness and nitrogen composition were provided, all on p-type InAs (Zn) substrates. Nitrogen compositions were determined using XRD also at Sheffield. However, it should be noted there is considerable uncertainty in these values, due to the unknown contribution of strain to any observed change in lattice constant. Relevant sample details are listed in Table 4.3, where values for $x$ assume 100% relaxation of the epilayer lattice.

<table>
<thead>
<tr>
<th>Sample Code</th>
<th>Epilayer Thickness (nm)</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2773</td>
<td>200</td>
<td>-</td>
</tr>
<tr>
<td>M2774</td>
<td>200</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M3008</td>
<td>300</td>
<td>0.001</td>
</tr>
<tr>
<td>M3012</td>
<td>400</td>
<td>0.004</td>
</tr>
<tr>
<td>M2775</td>
<td>300</td>
<td>0.005</td>
</tr>
<tr>
<td>M3009</td>
<td>300</td>
<td>0.007</td>
</tr>
<tr>
<td>M2808</td>
<td>300</td>
<td>0.01</td>
</tr>
<tr>
<td>M2806</td>
<td>300</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Table 4.3 The InAs$_{1-x}$N$_x$ samples investigated in this work. All samples were grown on p-type InAs (Zn) substrates. M2773 and M2774 are both control samples. M2774 was grown with the nitrogen source on, but the growth temperature was too high for incorporation.

It was anticipated that growing these epilayers on InAs substrates might preclude the possibility of performing absorption/transmission measurements on them. In fact, when attempts to do so were made, no transmission signal could be obtained whatsoever. However, given that such measurements are not expected to yield a true value for the bandgap anyway, this is not considered an important result. Of greater consequence, is the inability to perform accurate electrical measurements on the epilayers (e.g. Hall measurements), as one can expect to observe parallel conduction in the substrate and epilayer, the signals from each being difficult to differentiate.
Note there is some variation in epilayer thickness from sample to sample due to experimentation during growth. All are relatively thin layers though, which could lead to observation of PL emission from the substrate as well as the epilayer. However, for InAs the penetration depth, $1/\alpha$, at $\sim1064\text{nm}$ (the lasing wavelength in the forthcoming measurements) is $\sim500\text{nm}$. Therefore, as the samples are all around $\sim300\text{nm}$ thick the majority of emission (if not all) should be from the epilayer, and close enough to the surface that self-absorption is unimportant.

Two samples with InAs epilayers were grown as control samples. The second of these, M2774, was actually grown with the nitrogen plasma source on, though the temperature was too high for incorporation. This sample is, therefore, expected to be of nominally zero nitrogen composition (or $x<0.001$ as it appears in Table 4.3), although the presence of any interstitial nitrogen atoms may detrimentally affect its optical properties. From this point onwards, specific samples will be referred to by their sample codes.

### 4.4.2 Experimental Procedure

PL has been performed on each sample listed in Table 4.3 using the fast-scanning FTIR spectrometer described previously in section 1.1.3.3 to spectrally resolve the emission. The rest of the arrangement was essentially the same as that detailed in section 2.3.2. Samples were prepared $\sim5\times5\text{mm}$ in size, and mounted on copper plates with silver conductive paint. These were then placed in the same Cryophysics closed cycle He cryostat used for the previous GaSbN study. The window was of more concern on this occasion, as at low temperature the bandgap of InAs approximately coincides with the window's absorption band at $\sim2.8\mu\text{m}$. Also, its long wavelength cut-off at $\sim4.7\mu\text{m}$ is close to the PL emission wavelength anticipated from the highest composition samples. Once again it was deemed more valuable to be able to achieve temperatures below 77K, however, so that the samples' radiative efficiencies might be maximised. If window absorption was subsequently demonstrated to severely affect the spectra, the measurements could always be repeated later using the nitrogen bath cryostat with more suitable MIR windows. In an effort to make this determination early on, the sample of highest nitrogen composition, M2806, was the first to be measured. The window cut-off was thus found not to be a problem. A diode pumped
Nd:YAG laser was used as the PL excitation source, with maximum power output ~100mW at a wavelength of 1064nm. The spot size on the sample was ~1mm in diameter, giving a power density of around 10Wcm⁻². The maximum fringe modulation frequency due to FTIR mirror scanning was ~3.3kHz, so a laser modulation frequency of ~20kHz was chosen (i.e. >5 times 3.3 – see section 3.1.1). This was achieved using a mechanical chopper with vane widths ~1mm. In order to prevent simultaneous illumination of neighbouring vanes by the laser, it was necessary to focus the beam through the chopper (and subsequently re-collimate it) using lenses. The detector output was then processed by a lock-in amplifier, before being fed into the analogue-digital converter (ADC) of the spectrometer. This arrangement yielded a noticeable improvement in SNR over cw operation (i.e. direct input of chopper-free detector output into the spectrometer ADC). Higher order features are not an issue in FTIR spectroscopy, so order-sorting filters were not required. Furthermore, the excitation laser wavelength was far shorter than the minimum wavelength of interest, thereby allowing the measurements to be made without using any filters whatsoever.

Although PM techniques are not expected to aid in observing the true bandgap energy in heavily n-type material like InAsN (see section 4.2.3), it is thought PR might yet be useful for such measurements on binary samples, where any Moss-Burstein shift is likely to be small (i.e. M2773 and M2774). To that end, a ~10mW HeNe laser mechanically chopped at ~800Hz was used as the modulation source. Unfortunately, the fast scanning FTIR used for the PL in this study could not be used for PR due to the high fringe modulation frequency of its mirror (a more detailed explanation of which is provided in section 3.1.1). Also, at these wavelengths (i.e. 3-4μm) the step-scanning FTIR system was simply not stable enough to rival the higher specification grating system’s (see section 1.1.2.3) resolution or spectral quality, so the latter was used in these measurements (with 0.5μm blazed grating). No filter was used to screen laser scatter; only the GaSb wafer filter was available for these wavelengths, and it seemed to offer little benefit in test measurements. Instead, laser scatter was minimised as best as possible through careful alignment.

The same liquid nitrogen-cooled InSb detector was used for both PL and PR measurements. In the former case the detector’s matched pre-amp was used, while for
the latter the Stanford filter/amplifier was employed to enable extraction of the dc signal for $\Delta R/R$.

### 4.4.3 Results and Discussion

Results will once again be divided into two parts; those obtained on the binary InAs material (PL and PR), thereby providing validation of the experimental procedure and an experimental control, and the results obtained from the InAsN epilayers themselves (PL only).

The cryostat window response and the effects of atmospheric absorption have been corrected for in all PL spectra. All quoted/plotted PL transition energies may be assumed to be the peak energy less $k_B T/2$, unless otherwise stated. Where attempted, PR lineshape fitting was performed using the ‘marqu’ program designed and written by Dr. T. J. C. Hosea.

#### 4.4.3.1 Bulk InAs

##### 4.4.3.1.1 Photoluminescence

**M2773**

The PL signals from M2773 as a function of temperature is shown in Figure 4.12. At low temperature, the gross structure of the observed emission is in good agreement with the measurements of others.[32-35] Fitting the two peaks which are clearly resolved at 15K in Figure 4.12 using equations 2.10 and 2.11 has proven problematic, so differentiation of the spectrum has been performed to determine their transition energies. At 15K, the higher energy peak (labelled (a) in Figure 4.12), lies at $\sim 417.7 \text{meV}$, in reasonable agreement with the value of 416.7meV at 1.4K obtained by Lacroix et al on high-purity samples.[36] This is also the value recommended by Vurgaftman and Meyer for $E_g(T=0)$, though they also report literature values ranging from 0.41meV up to and above 0.42meV.[19] Based on the assignments of Lacroix et al, the lower energy peak (labelled (b) in Figure 4.12) situated at $\sim 402 \text{meV}$ at 15K is associated with recombination of a free electron and a hole bound to a shallow acceptor around 15meV above the valence band edge. It is likely there are several
other unresolved transitions within this feature, perhaps explaining at least in part why the two resolved peaks cannot be fitted using a single lineshape for each peak.

The high temperature PL emission from InAs has been reported previously to originate from band-to-band recombination.[33,34] However, fitting the PL emission from M2773 at 300K using equations 2.10 and 2.11, the transition energy is determined to be -339.3 meV, around -14 meV lower than the 300K bandgap energy suggested by Vurgaftman and Meyer (-353.3 meV).[19] However, the most recent works upon which that value is based are the PL measurements of Fang et al and Gong et al (i.e. the works which made the PL band-to-band transition assignment initially).[33,34] In particular, Fang et al assume the Boltzmann approximation is valid. Making a similar assumption in this case, differentiation of the 300K peak yields a transition energy -354.3 meV, in much closer agreement with the literature. In an effort to reconcile this study’s PL with the literature, from this point on those quoted for this sample will be evaluated assuming the Boltzmann approximation is valid, for the moment noting only that lineshape fitting using full Fermi-Dirac statistics does not yield the same energy.

Both the energy of the dominant transition in the PL from M2773, and the Varshni behaviour suggested by Vurgaftman and Meyer are plotted as a function of temperature in Figure 4.12. The two behaviours are thus seen to be in approximate agreement at low temperature (15K) and high temperature (250K and 300K), but show significant deviation (~5-10 meV) at intermediate temperatures. After close inspection of these results it is believed an explanation for this discrepancy can be offered. While previous works have assumed that it is band-to-band recombination that is responsible for the dominant/only peak at each temperature, here it appears that by 200K the dominant PL transition switches from band-to-band recombination, or peak (a), to a weak unidentified radiative transition (labelled (c) in Figure 4.12) lying close in energy to the band-to-band feature. Two pieces of evidence support a conclusion that the high temperature is, therefore, not from a band-to-band transition: firstly there is a pronounced kink in the PL transition’s temperature dependence between 150K and 200K, and secondly close inspection of Figure 4.12 indicates a rapid diminishing of (a) between 100K and 150K. The former observation is obviously unphysical if the same transition is being followed at all temperatures,
while to make the latter point clearer the band-to-band peak is annotated at 100K and 150K in Figure 4.12. Although very small, it can be seen that at 150K there is still a small remnant of (a) present in the spectrum, appearing to sit atop the broader feature, (c). The author suggests, therefore, that it is possible in previous temperature dependent PL studies peak assignments were confused at intermediate temperatures due to the indistinct nature of the cross-over between (a) and (c). Close inspection of the literature temperature dependent spectra as they appear in their respective publications does not rule out this interpretation.[34]

Figure 4.13 shows the excitation power dependence of the PL emission from M2773 at 15K. Laser power was decreased incrementally from ~100mW to ~1mW using neutral density (ND) filters. There is no discernable energy shift in the emission as the laser power is decreased, so the excitation does not appear to be causing significant localised heating of the lattice. A difference in the power dependence of the intensities of (a) and (b) can be seen however, with (a) exhibiting a stronger dependence than (b). This is consistent with the observations of others and is representative of their different origins.[35]
Bandgap Reduction in Dilute Nitride Materials Based on Narrow Gap Semiconductors

Figure 4.12 PL emission from M2773 between 15K and 300K. At low temperature (15K) and high temperature (200K-300K) there is good agreement between the suggested bandgap temperature dependence and the energy of peak (a) in this study.[19] However, there is significant deviation at intermediate temperatures. Unidentified feature (c) dominates the PL emission at high temperature (i.e. ≥150K). Intensity scaling factors for each spectrum are given.

Figure 4.13 Excitation power dependence of the PL emission from M2773 at 15K. The dashed black line passes through the centre of (a) at maximum excitation power (~100mW). There is no discernable energy shift as the power of the exciting laser is decreased (i.e. from bottom to top). Peak (a) has a stronger power dependence than peak (b), as noted by others previously.[15] Laser power was reduced using ND filters. Intensity scaling factors are indicated.
M2774

The PL from M2774 is shown in Figure 4.14. For this sample, (a) is considerably weaker relative to (b) than seen for M2773, probably due to the presence of interstitial nitrogen atoms. Most of the key discussion points made for M2773 apply to M2774 also, although the cross-over between (a) and (c) is rather less pronounced in this case. Whatever the origin of (c), its energy is apparently affected (slightly) by the introduction of interstitial nitrogen, as at high temperatures (200K-300K) the suggested literature values and derived PL transition energies for M2774 do not match as well as those for M2773.

Figure 4.14 The PL emission from M2774 between 15K and 300K. Here, (a) is considerably weaker relative to (b) than was seen in M2773, probably due to the presence of interstitial nitrogen atoms. The same key points made about M2773 apply to M2774 also, although the cross-over between (a) and (c) is less pronounced in this case. Intensity scaling factors are given.
4.4.3.1.2 Photomodulated Reflectance

M2773

The PR signal temperature dependence for M2773 is shown in Figure 4.15 (the region between 10 and 50K is shown expanded in Figure 4.16). To the author's knowledge, only one prior PM measurement on InAs exists in the literature.[37] In that work, Lin et al could measure PR only as low as 77K. Above that temperature, the lineshapes in Figure 4.15 are in broad agreement with that work's result on lightly doped $p$-type material ($p \sim 2 \times 10^{16} \text{cm}^{-3}$). To obtain critical point energies from their features, those authors noted that one particular lobe appeared to consistently match the peak position of the PL in parallel measurements. The extreme of that lobe was, therefore, chosen as the critical point energy (attributed to the $\Gamma$-point) at each temperature. However, there is considerable disagreement in the Varshni behaviours derived from those PL and PR measurements at high temperature, which the authors attribute to increased uncertainty in the PL peak position due to broadening of the lineshape. Furthermore, their PL Varshni behaviour shows substantial deviation from both Vurgaftman and Meyer's suggested behaviour, and the PL transition energies obtained from the emission of M2773 in this study.[19] However, it has been found that by subtracting $k_B T/2$ from the Varshni behaviour of Lin et al, good agreement with Vurgaftman and Meyer (and, therefore, M2773 at high temperature) can be obtained. This suggests Lin et al simply omitted that step in their analysis, which implies, unfortunately, their derived PR critical point energies might also be in error. This study's results, Vurgaftman and Meyer's suggested dependence, and the PL and PR of Lin et al are compared in Figure 4.17.

In Figure 4.15 the transition energy of (a) in the PL of M2773 is also plotted at each temperature. Attempts to fit the PR feature using modified Airy functions and TDFF lineshapes proved unsuccessful. However, it can be seen in Figure 4.15 that between 50K and 150K the PL energies consistently lie (approximately) at the centre of the PR feature. Above 150K this pattern is broken, thus reinforcing the argument that above 150K the PL emission from M2773 does not originate from the band edge. Hence, taking the extremum position of the central lobe of each PR feature (above 20K) as the critical point energies, a temperature dependence in excellent agreement with the PL of both M2773 and M2774 between 20K and 150K is found, with a smooth
evolution continuing above 150K (i.e. no kink). It is believed, therefore, that the PR is associated with the Γ-point (i.e. the band-to-band transitions) at all temperatures, making such measurements a more reliable means of determining the temperature dependent behaviour of the bandgap in (lightly-doped) bulk InAs than PL. The bandgap energies determined in this way are plotted in Figure 4.18.

In Figure 4.16 the region between 10K and 50K is shown expanded, with a 5K interval between sampled temperatures. At such temperatures (beyond the reach of Lin et al), a continuous evolution between one distinct lineshape at 50K and another entirely different one at 10K is clearly seen. It is believed this observation is consistent with a well documented effect involving screening of the exciton interaction by photo-excited carriers, which leads to the appearance of exciton-related features at low temperature.[38,39,40] The accompanying decrease in the amplitude of the FKO-like lineshape observed at high temperature is attributed to the freeze-out of carriers. As for bulk GaSb (see section 1.3.3.1.2), it is believed the strong PR signal observed from M2773 is a consequence of the electrostatic repulsion between the carriers at the p-type substrate/n-type epilayer interface. The resulting in-built field also leads to the formation of a depletion region in the neighbourhood of the layer interface. However, at low temperature carrier freeze-out reduces the field amplitude, increasing the effective width of the depletion region while simultaneously reducing the amplitude of the in-built field available for modulation. Therefore, the exciton-related and FKO-like features appear to exchange intensity as a function of temperature. The former does not exhibit FKO's because the field being modulated in that case (i.e. the electrostatic attraction between an electron and a hole) is small and, hence, so is $\hbar \Theta$. 

107
Figure 4.15 PR from M2773 between 15K and 250K. Note between 50 and 150K the PL transition energies coincide with the PR feature's centre. Taking that position as the PR critical point energy, a bandgap both consistent with the low temperature PL emission is obtained and smoothly varying over the full range is obtained. Intensity scaling factors are indicated. At 15K the feature is ≤2meV from the bandgap energy, consistent with excitonic origins (see Figure 4.16).[19]

Figure 4.16 PR from M2773 between 10K and 50K. The gradual evolution between two distinct lineshapes can clearly be seen. This is attributed to screening of the exciton interaction by photoexcited carriers at low temperature.[38,39,40]. The exciton binding energy in InAs is estimated to be ~1.1meV, close to $k_B T$ at 15K.[36]
Bandgap Reduction in Dilute Nitride Materials Based on Narrow Gap Semiconductors

Figure 4.17 Comparison of the derived PR critical point energies (believed to be associated with the Γ-point, or bandgap) for M2773, and various Varshni behaviours from literature sources. The PR and PL of Lin et al (including subtraction of $k_B T/2$) do not agree, although the latter’s PL is in good agreement with the behaviour suggested by Vurgaftman and Meyer.[19,37] The PR of Lin et al is close to the Varshni fit to this study’s PR, although the behaviours do have different slopes, crossing over at about ~250K. The PL transition energies for M2773 are also shown for $T \geq 200K$, illustrating the good agreement between the high temperature PL in all three studies.

Figure 4.18 Comparison of M2773 PR critical point energies between 15 and 250K (evaluated as the energy of the central extremum in the PR feature) and the PL transition energies of M2773 and M2774 between 15K and 150K (where (a) dominates). The PR energies are believed to offer a more reliable estimate of the bandgap energy above 150K, as the PL emission from both M2773 and M2774 (although in good agreement with the literature) appears to originate from elsewhere at those temperatures.
M2774

PR signals as a function of temperature for M2774 are shown in Figure 4.19. This measurement was performed first chronologically and, unfortunately, a slit width too large to fully resolve the lineshape at low temperature was inadvertently chosen. Time was not sufficient to repeat the measurement with narrower slits. However, a general form consistent with the PR of M2773 can still be seen, particularly at high temperature where the lineshape is broader. This provides some confidence in the repeatability of both the sample growth and characterisation technique. At 100K and 150K the PL peak energies from M2774 are again seen to coincide with the centre of the corresponding PR features. At 50K the lineshape is not fully resolved so any coincidence is obscured. Nevertheless, indications are that the correlation seen in the PR and PL from M2773 was not spurious. The amplitude of the PR signal from M2774 at 15K is around half that measured from M2773, however, probably indicating generally poorer optical quality.

Figure 4.19 PR from M2774 between 15K and 250K. Unfortunately, the slit widths in this experiment were chosen too large to fully resolve the feature at low temperature. Nevertheless, PR Lineshapes broadly similar to those observed from M2773 have been obtained, particularly for T>100K where the lineshapes are broadened. At 100K and 150K the transition energy of (a) in this sample’s PL coincide with the PR feature’s centre. The amplitude of the PR from M2774 at 15K is around half that from M2773. Intensity scaling factors are given.
4.4.3.2 InAs$_{1-x}$N$_x$ Epilayers

4.4.3.2.1 Photoluminescence

Figure 4.20 The low temperature (15K) PL emission from each InAsN sample, including control sample M2773 ($0 \leq x \leq 0.022$). A consistent unambiguous redshift is observed with increasing nitrogen composition.

An unambiguous incremental redshift has been observed in PL emission of InAs$_{1-x}$N$_x$ as a function of increasing nitrogen composition. This shift is most clearly seen at low temperature, as illustrated in Figure 4.20, where the emission spectra from each sample at 15K are shown.

The PL emission from all the InAsN samples exhibits just one resolvable peak at all temperatures. Even for nitrogen compositions as small as $x=0.001$, the features (a) and (b) common to InAs PL emission are absent, at least in their familiar form. No significant emission from the substrate could be discerned in the spectra. As it is anticipated that the InAsN samples are unintentionally doped highly $n$-type, application of the Boltzmann approximation is almost certainly invalid, particularly at high temperature. Indeed this was tested, with poor fits to the experimental features resulting. The 300K PL peaks have, therefore, been fit using equations 2.10 and 2.11.
It may be noted though, that the functional form they describe does not account for any non-parabolicity in the bands\textsuperscript{13}, presenting a possible source of error when applied to InAs\textsubscript{1-x}N\textsubscript{x}. However, as was demonstrated in section 2.7.1, the enhancement in the effective mass of the $E_{\text{sub}}$ subband is actually predicted to be quite gradual in InAsN. For example, at $x=0.02$ (representative of the highest composition studied) Figure 4.4 shows the effective mass ratio of $E_{\text{sub}}$ remains close to unity (i.e. $m_{\text{M}}/m_{\text{M}}<1.3$) as high as $\sim0.3\text{eV}$ above the band edge, an order of magnitude greater than the anticipated bandgap reduction of $\sim0.06\text{eV}$. Hence, at such compositions band parabolicity should remain a reasonable approximation in the region of the conduction band where carriers partaking in radiative recombination are concentrated. The transition (or bandgap) energies obtained from the fitting procedure at 300K are plotted in Figure 4.12, along with the results obtained by other groups.[12,13,24] The fit of BAC model (see section 4.2.1) to the PL transition energies is also shown.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4_21.png}
\caption{The PL transition (or bandgap) energies of the InAsN samples studied in this work as a function of nitrogen concentration at 300K (red dots), and a BAC fit to those energies (black line). The binary bandgap (i.e. the $y$-intercept) is the value obtained from PR measurements on InAs control sample M2773. Also shown in other colours are the body of existing data from absorption measurements by other authors. Note the results of Naoi et al and Kuroda et al are quoted as approximate due to uncertainties when reading the values from graphs in their publication form.[12,13,24]}
\end{figure}

\textsuperscript{13} Note the intrinsic non-parabolicity of InAs itself has been neglected throughout.
The resulting BAC parameter values are listed in Table 4.4, along with the initialising parameters provided by O'Reilly et al based on a theoretical tight-binding calculation.[11] The value of $E_M^0$ is the 300K bandgap of InAs control sample M2733, as measured using PR earlier. That parameter was fixed during the fit. Of immediate note, good agreement between the initialising (theoretical) and fit (experimental) parameter values. This can also be seen in Figure 4.21, where the line corresponding to the theoretical BAC prediction (based on the initialising parameters) and the BAC experimental fit line compare well. However, the fact that the sample concentrations measured tend to be clustered at low values ($x \leq 0.01$) gives a great deal of unwarranted weight to the $x=0.022$ point in the fit. Ideally, more sample concentrations intermediate between $x=0.01$ and $x=0.022$ should be measured. This is perhaps an area for future work. In terms of agreement with prior studies, the values of $\beta = C_{NM}$ and $E_N^0$ in Table 4.4 are considerably smaller than those of $\beta = 1.68$ and $E_N^0 = 1.48$ determined by Shih et al.[12] An overall redshift of 60meV has been found for the highest composition ($x=0.022$) sample investigated, corresponding to reduction of ~27meV per percentage nitrogen added.

<table>
<thead>
<tr>
<th>BAC Model Parameter</th>
<th>Fit Value</th>
<th>Initialising Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$1.34 \pm 0.40eV$</td>
<td>$1.30eV$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$1.40 \pm 0.13eV$</td>
<td>$1.30eV$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$2.00eV$</td>
<td>$2.00eV$</td>
</tr>
<tr>
<td>$E_N^0$</td>
<td>$1.30 \pm 0.23eV$</td>
<td>$1.36eV$</td>
</tr>
<tr>
<td>$E_M^0$</td>
<td>$-0.358eV$</td>
<td>$0.358eV$</td>
</tr>
</tbody>
</table>

Table 4.4 Parameters resulting from a fit of the BAC model to the 300K PL transition energies. The fit was extremely insensitive to the value of $\gamma$, so no error is quoted (it is very large though). The parameter $E_M^0$ (the measured bandgap of control sample M2773) was fixed during the fit. Initialising parameters are from tight-binding calculations by O'Reilly et al.[11]

The 300K Fermi levels of for each sample, as calculated based on the quasi-Fermi energy fit values (see section 2.2.1), are plotted in Figure 4.22. For higher composition samples the Fermi level is indicated to lie far above the band edge. Its approximate position around ~90meV above the band edge for $x=0.022$ is roughly
consistent with the (multiple) equivalent values quoted by Shih et al.[12]

![Diagram of Fermi level and conduction band minimum vs nitrogen composition](image)

**Figure 4.22** The extracted InAsN Fermi level for nitrogen concentration between 0.1% and 2.2%. Values were calculated from the quasi-Fermi levels, obtained from a fit of equations 2.10 and 2.11 to the experimental PL peaks of each sample. Aside from the spurious point at 0.7% (M3009), the conduction band carrier concentrations of the measured samples are degenerate at room temperature. For 2.2% nitrogen the Fermi energy is ~90 meV above the band edge.

Attempts were also made to fit the lower temperature PL spectra in a similar way to the 300K peaks. However, although a visibly reasonable fit could be obtained for many such spectra, as one decreased the temperature the fit parameter values (e.g. quasi-Fermi levels, transition energy, etc.) became increasingly unphysical. This is most likely due to the presence of multiple recombination paths at lower temperatures, arising from the recombination of carriers trapped in localised nitrogen cluster states just below the band edge (i.e. the InAsN equivalent of those states predicted by LCINS calculations for GaAsN).[16] The lineshape at such temperatures is, therefore, believed to be a composite of several unresolved PL peaks. In support of this hypothesis, the unphysical behaviour of the lineshape fit parameters was found to be significantly worse in higher concentration samples (e.g. M2808 and M2806), with its onset seen to occur at higher temperature.

It has been observed previously in nitrogen containing alloys, that the multiple PL peaks arising from this electron trapping mechanism result in an anomalous temperature dependence of the composite peak as a whole.[41,42,43] That is, rather than exhibiting a Varshni-like behaviour (see appendix, section 9.1), a peak may
initially appear to move toward higher energy as the temperature is increased from low values. This is due to ionisation of the nitrogen cluster-related states lying just below the bandgap, leading to an exchange in intensity between transitions originating from such lower energy nitrogen states and those from the \( E_{\text{sub}} \) band states. At some temperature all such states are ionised, and the non-Varshni behaviour is pacified. The peak is then observed to change its direction of movement as a function of temperature (i.e. it shifts toward lower energies).

To check for that signature of dilute nitride alloys, the PL peak energy, \( E_{\text{max}} \), as a function temperature have been determined by differentiation (note in this case \( k_{\text{B}}T/2 \) was not subtracted). The temperature dependences of the PL emission for several representative (higher composition) samples are shown in Figure 4.23 to Figure 4.26, where the derived peak energies are also plotted. All the samples presented therein are seen to exhibit the expected anomalous temperature dependence, albeit to different degrees. In particular, the variation of M2808 \((x=0.01)\) shows the most extreme behaviour, with a well-defined reversal in its variation at about 150K. However, contrary to expectation, the higher composition sample M2806 \((x=0.022)\) exhibits a much weaker anomalous behaviour. This weaker behaviour is attributed to certain nitrogen cluster-related states (perhaps those associated with pairs) gaining sufficient density that their interaction with the \( E \) band provides them with significant extended (or \(|k\rangle\)) character, thereby leading to the formation a band of states.[16] This band thus defines a new band edge, with only the sparsely occurring nitrogen triplet cluster states found lower in energy giving rise to the weak anomalous temperature dependence observed in M2806. Therefore, one might expect the anomalous behaviour to become stronger once again for \( x>0.022 \), as the nitrogen triplet-related clusters become more numerous. It is hoped that this hypothesis can be tested in the future if higher composition samples become available for such measurements.
Figure 4.23 PL from InAs$_{0.993}$N$_{0.007}$ sample M2775 between 15K and 300K. The peak position, $E_{\text{max}}$ (without subtracting $k_B T/2$) is indicated at each temperature (blue dots). The anomalous temperature dependent variation of the peak can be seen at low temperatures (i.e. peak position initially moves to higher energy with increasing temperature). This implies the more observed emission peak is comprised of more than one radiative path. The 300K transition energy plotted in Figure 4.21 is indicated by an arrow. Intensity scaling factors are indicated.

Figure 4.24 PL from InAs$_{0.993}$N$_{0.007}$ sample M3009 between 15K and 300K. The peak position, $E_{\text{max}}$ (without subtracting $k_B T/2$) is indicated at each temperature (blue dots). The anomalous temperature dependent variation of the peak is slightly less pronounced than in M2775 (see Figure 4.23), though still visible. The 300K transition energy plotted in Figure 4.21 is indicated by an arrow. Intensity scaling factors are indicated.
Figure 4.25 PL from \( \text{InAs}_{0.990}\text{N}_{0.010} \) sample M2808 between 15K and 300K. The peak position, \( E_{\text{max}} \) (without subtracting \( k_{B}T/2 \)) is indicated at each temperature (blue dots). The anomalous temperature dependent variation of the peak is strongest in this sample, with a pronounced reversal at around 150K. The 300K transition energy plotted in Figure 4.21 is indicated by an arrow. Intensity scaling factors are indicated.

Figure 4.26 PL from \( \text{InAs}_{0.988}\text{N}_{0.022} \) sample M2806 between 15K and 300K. The peak position, \( E_{\text{max}} \) (without subtracting \( k_{B}T/2 \)) is indicated at each temperature (blue dots). In contrast to M2808 (see Figure 4.25), only a weak anomalous temperature dependence is observed. Hence, at this composition the cluster-related states are believed to have gained sufficient strength to form a band, redefining the band edge. The 300K transition energy plotted in Figure 4.21 is indicated by an arrow. Intensity scaling factors are indicated.
4.4.4 Conclusions and Future Work

The PL emission from InAsN has been investigated as a function of nitrogen concentration and temperature. An unambiguous bandgap reduction has been observed at low and high temperature. Fitting the experimental data using a lineshape incorporating Fermi-Dirac occupation functions, the composition dependence of the derived transition energies is in good agreement with BAC model predictions of the material’s bandgap. An overall redshift of 60meV has been found for a composition of \( x = 0.022 \), corresponding to a reduction of \( \sim 27\text{meV} \) per percentage nitrogen. In the same sample, the Fermi level has been estimated to lie \( \sim 90\text{meV} \) above the conduction band edge for \( x = 0.022 \).

The temperature dependence of the PL peak energy for higher nitrogen compositions also appears to exhibit an anomalous behaviour, confirming the presence of a classic signature of dilute nitride alloys. The pacification of the effect for the highest nitrogen composition sample (\( x = 0.022 \)) is believed to be consistent with the LCINS model.

In addition, experiments on the bulk InAs (provided for as control samples for the dilute nitride work) indicate the room temperature PL emission does not originate from a band-to-band transition. This is in contradiction to accepted understanding. The accompanying PR measurements yield a more convincingly accurate temperature dependence at high temperature, while remaining entirely consistent with the low temperature PL.

Looking to the future, it is hoped to be able measure more InAsN samples in order to fill in some of the composition gaps (e.g. \( 0.01 < x < 0.022 \) and \( x > 0.22 \)) left by this study. Also further studies of InAs using both PR and PL are necessary to determine whether the signals obtained in this work are indicative of the material generally, or a peculiarity of the samples measured in this study.
4.5 References


5 Time-Resolved Measurements of Spin Relaxation in Bulk Narrow Gap Semiconductors

5.1 Introduction

5.1.1 The Spin-FET: A Spin-Based Semiconductor Device

The key advantages of electronic devices which take advantage of the electron spin have been discussed previously in section 1.3.1, where some examples of current and future devices were also mentioned. Also, the need for a deeper understanding of the physics governing spin transport and relaxation was stated. In order to add some further context to such issues it is instructive to examine one of the proposed devices in greater detail. To this end, consider the spin-FET originally proposed in 1990 by Datta and Das.[1,2] The basic structure of their device consists of a 1D semiconductor channel terminated at each end by a ferromagnetic metal (or semiconductor) source and drain respectively. The source (spin-injector) and drain (spin detector) are arranged with their magnetisations aligned parallel. A gate electrode situated above the channel is then used to apply a voltage $V$ across the channel. Due to spin-orbit coupling (a relativistic effect), electrons transiting the channel experience a unidirectional magnetic field proportional to $\nabla V$, about which their spins precess at a frequency proportional to the magnitude of the field (see appendix, section 9.5). If the period of that precession is very much longer than an electrons time of flight across the channel (i.e. the source to drain transit time) they will reach the drain with their
spins still aligned (approximately) parallel to its magnetisation, leading to a large current in any external circuit (following from the GMR effect – see section 1.2.1). However, if the alignment of their spins is reversed during flight (anti-parallel to the ferromagnet magnetisation) a smaller current passes through the device. To achieve the required tunability of the precession frequency the gate voltage can be varied, changing $\nabla V$ and, hence, the magnetic field strength. Thus it can be seen that even this most basic design for a spin-FET depends on the physics governing spin injection and extraction across ferromagnetic/semiconducting boundaries, and spin decoherence/relaxation (i.e. the spins must retain their polarisation during passage between source and drain). Research in these areas of spin-physics is at an early stage, so the physical processes involved remain relatively poorly understood. Nevertheless, a great deal of investment is currently being made to try and further that understanding.

A key advantage of the spin-FET over a regular FET is that the energy required to flip an electron’s spin (~100µeV) is considerably less than the kinetic energy required to physically move the electron (~10’s meV), potentially making the spin-FET a more efficient device. Furthermore, spin-precession times (~ns) can be far shorter than the time required to expel electrons from the conducting channel of a regular FET (~µs), leading also to faster switching times. A hypothetical spin-FET could, therefore, be integrated into the circuitry of a computer’s CPU (Central Processing Unit) to produce a chip which is (in principle) faster (i.e. higher clock speeds) whilst simultaneously more power efficient than the present state of the art.[2,27]

5.2 Theory

5.2.1 Cubic Spin-Splitting in Bulk III-V Semiconductors

Spin-orbit coupling’s most apparent influence on the bandstructure of common semiconductor materials such as III-V alloys, is the splitting of the highest valence band states into the heavy hole, light hole, and split-off (or spin-orbit) bands. Additionally, it can also lead to considerably more subtle effects which are often ignored in discussions of gross bandstructure. These are spin-splittings proportional to odd powers of the wavevector, $\mathbf{k}$; those proportional to $|\mathbf{k}|$ are referred to as ‘linear’
splittings, while those proportional to $|k|^3$ are referred to as ‘cubic’ splittings.[3,4,5] Higher powers are possible, but their magnitudes are usually far smaller. In general, linear and cubic splittings are not mutually exclusive, the two usually occurring simultaneously. However, in bulk semiconductors the considerably weaker dependence of linear splittings on $|k|$ means the effect of cubic splittings tend to dominate, particularly for $|k|>0$. Therefore, the influence of linear splittings will be omitted from the discussions and analyses of this chapter, though they will be treated in some detail in chapter 6.

Many texts introduce the concept of spin-orbit coupling with reference to a simple semi-classical example such as the Bohr model of the hydrogenic atom.[6] Within this model the orbiting atomic electron is treated as a current loop (or magnetic dipole) moving through the electrostatic field of the proton with quantised angular momentum, $L$. Switching to the rest frame of the electron, the proton is observed orbiting the electron, thus generating a magnetic field, $B$, at the site of the electron. This field can then interact with the electron’s magnetic dipole moment, $\mu_s$, associated with its spin, $S$, lifting the degeneracy of the spin-eigenstates. The spin-orbit Hamiltonian describing the interaction, $\mathcal{H}_{so}$, may be written as follows:

$$\mathcal{H}_{so} = \lambda L \cdot S$$  \hspace{1cm} (5.1)

where $B$ and $L$, and $\mu_s$ and $S$ are related via constants of proportionality contained within $\lambda$.

Applying the full relativistic formulation of quantum mechanics to the hydrogenic atom, where the electron has orbital angular momentum dependent on orbital quantum number, $l$, it may be shown that the expression for the spin-orbit interaction may be reduced to a form identical to that of equation 5.1 (which followed from entirely classically-based arguments) as follows:[7]

$$\mathcal{H}_{so} = \frac{\hbar^2}{4m_e^2c^2} \frac{1}{r} \frac{dV}{dr} (r \times p) \sigma$$  \hspace{1cm} (5.2)

$$= \lambda' L \cdot S$$  \hspace{1cm} (5.3)
where the constant \( \lambda' = (dV/dr)/(2mc^2) \) and spherical symmetry is assumed.

Therefore, spin-orbit coupling may be expressed in terms of the interaction of \( S \) with an appropriately chosen ‘effective’ magnetic field, \( B_{\text{eff}} \), by combining equation 5.1 and equation 5.3:

\[
\mathcal{H}_{\text{so}} = \lambda'' B_{\text{eff}} \cdot S \tag{5.4}
\]

where for this hydrogenic atom example \( B_{\text{eff}} \parallel \mathbf{L} \), and \( \lambda'' \neq \lambda' \) still represents the fundamental strength of the interaction. In this particular case, there is actually little benefit in expressing the interaction in this way. However, it should be apparent that by eliminating the explicit dependence of \( \mathcal{H}_{\text{so}} \) on \( \mathbf{L} \) the expression has been made far more general. It is, therefore, a far more useful expression for any case where one is not dealing with spherical symmetry. Generally, the form of \( B_{\text{eff}} \) varies depending on the specifics of the spin-orbit interaction in question, though it may be said that it always characterises the interaction’s dependence on a relative state of motion. The constant \( \lambda'' \) in equation 5.4 contains any constants of proportionality, which for an electron in the conduction band includes material specific parameters such as the bandgap energy and conduction band effective mass. It must be emphasised at this point, that although \( B_{\text{eff}}(k) \) is not a true classical field, it is still far more than just a convenient choice of nomenclature. It is expected that any spins subject to a spin-orbit interaction exhibit the same field dependent behaviours discussed in the appendix, section 9.5 (e.g. precession, superposition/mixing etc.).

As alluded to above, the symmetry of a semiconductor’s crystal lattice plays a pivotal role in determining the functional form of \( B_{\text{eff}}(k) \). The symmetry of a crystal is mapped directly into reciprocal space, with the symmetry of the Brillouin zone (in ‘reciprocal’ space) reflecting the symmetry of the crystal’s unit cell (in ‘real’ space). Most III-V semiconductors are based on the zinc-blende crystal lattice. This consists of an fcc (face-centred cubic) structure, with each lattice site associated with two atoms of different atomic species, offset from one another along the body diagonal of

---

14 The use of the word ‘effective’ is intended to emphasise that this is not a physically ‘real’ classical field.
the cube by a quarter of the unit cell dimension, \( a \). Hence, the primitive cell of a zinc-blende lattice contains two different atoms. The diamond lattice of elemental group IV semiconductors such as Si and Ge is identical to that of zinc-blende, only it is comprised of one only one atomic species. Thus the primitive cell of a diamond lattice contains two identical atoms. The zinc-blende and diamond lattice structures are illustrated in Figure 5.1.

![Diamond Lattice](image1)

![Zinc-Blende Lattice](image2)

**Figure 5.1** The structure of the diamond (LHS) and zinc-blende lattices (RHS). Two unit cells of dimension \( a \) are shown in each case. Bonds to nearest neighbour atoms are shown in green. the zinc-blende lattice consists of two interpenetrating fcc lattices, offset from one another by \( a(1/4, 1/4, 1/4) \). Each fcc lattice is comprised of a different atomic species (i.e. one group III atoms, and the other group V). The zinc-blende primitive cell is the smallest volume containing a lattice site from each component fcc lattice (i.e. an atom of each species). The diamond lattice has the same basic structure, except the atoms of both fcc lattices are of the same type (i.e. group IV), and the primitive cell thus contains two identical atoms.

In terms of relevance to this work, the most important difference between the two lattice structures is that the diamond lattice is invariant under inversion about a point at the centre of the bond between atoms in the primitive cell, whereas the zinc-blende lattice is not. Therefore, the net potential, \( V \), of the atoms in the diamond lattice’s primitive cell must also be inversion symmetric about the centre of the atomic bond. This implies that \( V \) is a purely even function within the primitive cell, as illustrated in Figure 5.2. Noting that an electron’s Bloch function is related to the Fourier transform of \( V \), one can deduce that the Bloch function must also be a purely even function, necessitating there be no dependence on odd powers of \( |k| \) (as such terms would be odd functions) in the electron Hamiltonian. However, the electronegativity of the bond in a zinc-blende lattice’s primitive cell leads to a net potential lacking inversion
symmetry about the centre of the bond, as shown in Figure 5.2. Within the primitive cell \( V \) must, therefore, contain odd Fourier components, which under Fourier transformation give rise to certain odd components in the Bloch functions, leading to odd-powered \(|k|\) in the electron Hamiltonian. These may include cubic (and linear) terms. One can, therefore, conclude that, unlike for the diamond lattice, cubic (and linear) spin-orbit coupling can (in principle) be present in a material based on the zinc-blende lattice structure.

![Figure 5.2 Schematics illustrating the symmetry properties of the potential (as experienced by an electron) within the primitive cell for the diamond (LHS) and zinc-blende (RHS) lattices.](image)

The spin-orbit interaction directly induces splittings in the p-like valence band orbitals, with the split-off energy, \( \Delta_{\text{so}} \), proportional to the size of the interaction. The magnitude of the spin-orbit splitting manifest in the conduction band is dependent on its degree of mixing with the spin-split valence band, particularly the light-hole band. As such, mixing becomes stronger the closer the states of each band are in energy, the spin-splitting of the conduction band exhibits an inverse dependence on the bandgap and conduction band mass, \( m_c \) (assumed to be approximately equal to the light hole mass). Hence, NGSs generally exhibit a larger conduction band spin-splitting than WGSs. Additionally, NGSs tend to be formed from atoms with higher atomic numbers than WGSs, so the magnitude of \( V \) may also be expected to be correspondingly larger in such materials, thereby increasing the magnitude of \( B_{\text{eff}}(k) \).
and, hence, the spin splitting further still.\(^\text{15}\)

In the case of electrons in the conduction band of a bulk zinc-blende material, the form of \(B_{\text{eff}}(k)\) may be shown to be:[3]

\[
B_{\text{eff}}(k) = \eta \left[ k_x (k_y^2 - k_z^2), k_y (k_z^2 - k_x^2), k_z (k_x^2 - k_y^2) \right] (5.5)
\]

\[
= \eta \kappa (5.6)
\]

where \(\eta = \hbar^2 (2m_e E_g e^2)^{-1/2}, \) \(A\) being a dimensionless constant representing the strength of the coupling interaction. When discussing spin relaxation times, \(B_{\text{eff}}(k)\) is usually expressed in terms of the precession frequencies, \(\Omega(k)\), of the electrons subject to it, which is related to the field as follows:[2,8,9]

\[
\Omega(k) = \frac{e}{m_e} B_{\text{eff}}(k) (5.7)
\]

\[
= A \hbar^2 \left( 2m_e^3 E_g \right)^{-1/2} \kappa (5.8)
\]

Note equation 5.7 may be arrived at following the analysis in the appendix, section 9.5. Multiplying equation 5.8 by \(\hbar\), an expression for the energy difference between spin-eigenstates may be obtained. The resulting material specific constant, \(A \hbar^2 (2m_e^3 E_g)^{-1/2}\), collectively has units of meV cm\(^3\), and gives the energy splitting of the spin-eigenstates per unit wavenumber cubed. This splitting ‘coefficient’ is often referred to as the Bulk (or 3D) Dresselhaus term.[3] The constant \(A\) is given by:[10]

\[
A \approx \frac{4 \delta}{\sqrt{3} - \delta} \frac{m_e}{m_0}, \quad \delta = \frac{\Delta_{\text{so}}}{(E_g + \Delta_{\text{so}})} (5.9)
\]

Inspection of equation 5.5 reveals that the magnitude of the spin-splitting is not only proportional to \(|k|^3\), but also highly directional. This is easily demonstrated by substituting the values for certain directions into equation 5.5 (e.g. [110] corresponds to \(k_x=1, k_y=1\) and \(k_z=0\)). Along <100> and <111> it can, therefore, be shown that there

\(^{15}\) These represent significant advantages to using NGSs in spintronic applications.
is no spin-orbit splitting, while along <110> it is maximised. Intermediate directions yield splitting magnitudes somewhere between these extremes. For example, along <301> the splitting is exactly one quarter of that observed along <110>. The orientation of \( \mathbf{B}_{\text{eff}} \) and, therefore, that of the spin-eigenstates, can be found using a similar methodology. For example, for \( \mathbf{k}||[110] \) equation 5.5 reads \( \mathbf{B}_{\text{eff}}=\begin{bmatrix} 1 \times (1-0), 1 \times (0-1), 0 \times (1-1) \end{bmatrix}=\begin{bmatrix} 1, -1, 0 \end{bmatrix} \), informing that the field is aligned with the [1-10] direction, perpendicular to \( \mathbf{k} \). Figure 5.3 illustrates a constant energy surface for a spin-split conduction band in the \( xy \)-plane, with the orientation of \( \mathbf{B}_{\text{eff}} \) along <110> indicated by arrows. To obtain the corresponding figures for the two orthogonal planes, \( yz \) and \( xz \), one need only perform cyclic permutation of the coordinates, as the band dispersions are identical.

![Figure 5.3](image)

**Figure 5.3** Constant energy surfaces of conduction band states split as a result of cubic terms in the crystal Hamiltonian. The orientations of \( \mathbf{B}_{\text{eff}} \) along <110> is indicated by arrows. For any crystallographic direction \( \mathbf{B}_{\text{eff}} \) is always perpendicular to \( \mathbf{k} \).

Figure 5.4a shows a two dimensional \( E-\mathbf{k} \) diagram of the same bandstructure illustrated in Figure 5.3, with Figure 5.4b and Figure 5.4c showing 1D slices through Figure 5.4a. In these examples, parabolic bands have been assumed under zero-field conditions, with bandstructure parameters suitable for bulk InAs used in equations 5.8 and 5.9.[24] It should be noted that the magnitude of the splitting in Figure 5.3 and Figure 5.4 has been exaggerated somewhat in the interests of clarity. In this example, approximately \(~100\text{meV} \) above the conduction band edge, a realistic value for the splitting along <110> is of the order of \(~3.5\text{meV} \), corresponding to a value for the Dresselhaus term of \(~1.9\times10^{-22}\text{eVcm}^3 \).
5.2.2 Spin Relaxation Mechanisms

In bulk semiconductors, it is believed there are three main spin relaxation paths, the physics of which are described by the Bir-Aronov-Pikus (BAP)[11], D’yakonov-Perel (DP)[8,9], and Elliott-Yafet (EY)[12,13] mechanisms respectively. The BAP mechanism is believed only to be dominant in p-type WGSs and is, therefore, not of direct relevance to this work. Of the remaining two mechanisms, their respective dependences of the spin relaxation rate on temperature and mobility provide a means of identifying which is dominant in a given material.

5.2.2.1 The D’yakonov-Perel Mechanism

The DP mechanism is reliant on the precession of spins about the localised fields which occur when a crystal lattice lacks a centre of inversion symmetry. In general any population of spin-polarised electrons must occupy states with a distribution of $k$, regardless of the method used to create them (e.g. electrical injection, photo-excitation), due to the Paul Exclusion Principle. This leads to a distribution of spin-
Time-Resolved Measurements of Spin Relaxation in Bulk Narrow Gap Semiconductors

precession frequencies, $\Omega$. In considering such a system one can conceive two key regimes; momentum relaxation (or scattering) rate, $1/\tau_p$, much larger than the distribution averaged precession frequency $\Omega_{av}$, or $\Omega_{av}\tau_p \leq 1$, and the reverse case of $\Omega_{av}\tau_p \geq 1$. In the latter case, dephasing initially proceeds reversibly due to the different precession frequencies present in the distribution. The spin is only lost irreversibly when the phase is randomised by momentum scattering collisions. When referring to the DP mechanism, it is usually the former regime (when $\Omega_{av}\tau_p \leq 1$) that is of relevance, so further discussion will focus on that case.[2]

From the point of view of a single electron spin, momentum scattering collisions amount to a randomly fluctuating $B_{eff}(k)$, both in terms of magnitude and direction. Hence, a spin precessing about such a field follows a random walk. The average period between steps in the spin’s random walk is given by $\tau_p$, and during that time the phase of a spin progresses by $\delta \varphi = \Omega_{av}\tau_p$. On a 2D phasor diagram, where the $y$-axis represents the $z$-component of the spin, there is an equal probability of the spin’s precession axis rotating in the clockwise or anti-clockwise sense. Therefore, after some time $t$, equivalent to $t/\tau_p$ steps, the phase of a single spin will have rotated through an angle $\varphi(t) \approx \delta \varphi (t/\tau_p)^{1/2}$. This is a standard result for any 2D random walk. Conventionally, the associated spin relaxation time, $\tau_s$, is defined as the time at which $\varphi(t)=1$, yielding the following expression:[2]

$$\frac{1}{\tau_s} = \Omega_{av}^2 \tau_p$$

(5.10)

Hence, the spin relaxation rate is inversely proportional to the momentum relaxation rate, suggesting the more scattering an electron undergoes, the longer its spin lifetime will be. This rather counter-intuitive result is closely related to a common dephasing phenomenon encountered in NMR known as motional (or dynamical) narrowing; essentially an inhibition of phase change due to a randomly applied ‘force’.[14]

Combining equations 5.8 and 5.10 yields the following expression for the spin relaxation rate in a regime of degenerate $n$-type doping:[10]

---

15 In NMR nuclear spins dephasing in this way can be re-phased using what is known as a 180° pulse, essentially a $\pi$ phase shift in their precession, to obtain a ‘spin-echo’.

134
Time-Resolved Measurements of Spin Relaxation in Bulk Narrow Gap Semiconductors

\[
\frac{1}{\tau_s} = \Gamma A^2 \frac{E_F^3}{\hbar^2 E_g} \tau_p \quad (5.11)
\]

Where \( \Gamma \) is a dimensionless collection of constants expressing the effectiveness of the dominant momentum scattering process at randomising the spin precession frequency, \( E_F \) is the Fermi energy, and \( A \) is the constant defined by equation 5.9. Because \( E_F \) varies only slowly with temperature, the temperature dependence of the spin lifetime in this case is dominated by that of the \( \tau_p \) and, hence, the mobility of the material in question.

Likewise, in the non-degenerate regime the spin relaxation rate is given by:[10]

\[
\frac{1}{\tau_s} = \Gamma' A^2 \frac{(k_B T)^3}{\hbar^2 E_g} \tau_p \quad (5.12)
\]

Where \( \Gamma' \neq \Gamma \) generally, and \( k_B T \) is the thermal energy of an electron (\( E_F \) is below the band edge). In this case, \( k_B T \) is a more rapidly varying function of temperature than \( E_F \) and, therefore, \( (k_B T)^3 \) dominates the temperature dependence of the spin lifetime.

Hence, in the non-degenerate regime one might expect to find a spin lifetime with a strong dependence on temperature, whereas in the degenerate regime a relatively weak dependence is suggested. Furthermore, as the DP mechanism effectively suppresses spin dephasing processes, one can expect the associated spin lifetime to be shortest when \( \Omega_{av} \tau_p \gg 1 \). Therefore, the DP mechanism is most likely the dominant relaxation mechanism (i.e. the spin lifetime is shortest) when the cubic splitting of the conduction band is large (i.e. \( B_{eff}(k) \) and \( \Omega_{av} \) are large), or \( \tau_p \) is very long (i.e. high mobility), as in intrinsic or lightly n-type doped narrow gap III-V semiconductors.

5.2.2.2 The Elliott-Yafet Mechanism

Like the DP mechanism, the EY mechanism requires the presence of spin-orbit coupling in the material bandstructure to proceed. However, in contrast to DP which relies on the anisotropic nature of cubic spin-splitting of the conduction band to
randomise the spin-precession axis during scattering, any spin-orbit interaction (anisotropic or otherwise) is sufficient for the EY mechanism. The EY mechanism can, therefore, provide a spin relaxation path in materials beyond those lacking inversion symmetry.[2] In the presence of any spin-orbit interaction (or $B_{\text{eff}}$), the Bloch states of an electron in the valence band are, generally speaking, not spin-eigenstates, but a superposition of the two (see appendix, section 9.5). During spin-conserving momentum scattering processes (i.e. those that operate only the spatial component of the electron's wavefunction), there is, therefore, a finite probability of the electron's spin being flipped. On average, such processes serve to randomise any spin-polarisation. Hence, it is an induced linkage between the spatial and spin components of the electron wavefunction in the presence of spin-orbit coupling that is fundamentally responsible for the EY relaxation mechanism.

The spin relaxation rate due to the EY mechanism in a regime of degenerate regime $n$-type doping is given by:[10]

$$\frac{1}{\tau_s} = \Lambda \delta^2 \left( \frac{E_F}{E_g} \right)^2 \left( \frac{1 - \delta/2}{1 - \delta/3} \right)^2 \frac{1}{\tau_p} \quad (5.13)$$

where $\Lambda$ is a dimensionless collection constants describing the effectiveness of the dominant scattering process on spin-flipping (not generally the same as $\Gamma$ or $\Gamma'$), and $\delta$ is defined by equation 5.9. Note here the reverse dependence of the EY mechanism on scattering rate as compared to DP, with the spin relaxation rate directly proportional to the momentum relaxation rate. Additionally (and as stated previously with respect to the DP mechanism), in the degenerate regime $E_F$ is a relatively weak function of temperature, so it can be expected that the material's mobility dominates the temperature dependence of the spin lifetime.

Likewise, in the non-degenerate regime, the spin relaxation rate given by:[10]

$$\frac{1}{\tau_s} = \Lambda' \delta^2 \left( \frac{k_B T}{E_g} \right)^2 \left( \frac{1 - \delta/2}{1 - \delta/3} \right)^2 \frac{1}{\tau_p} \quad (5.14)$$
where $\Lambda \neq \Lambda$. Here, the spin relaxation time is proportional to only the square of $k_B T$, rather than the cube as in the DP mechanism. Hence, although in this regime $(k_B T)^2$ dominates the temperature dependence, a lower sensitivity of the EY mechanism to temperature might be expected than for the DP mechanism.

Because the rate of spin flipping in the EY mechanism is proportional to the degree of mixing between spin states, the process efficiency scales with $\Delta_{so}$ (i.e. the strength of the spin-orbit coupling in the valence band) and inversely with $E_g$ (i.e. the degree of mixing between conduction and valence band). Therefore, it is most likely to be the dominant relaxation mechanism in materials where $\Delta_{so}$ is large and $E_g$ is small, or $\tau_p$ is very short (i.e. low mobility), as in heavily $n$-type doped narrow gap III-V semiconductors. It also may be noted that in the non-degenerate regime the different dependences of the EY and DP mechanisms on temperature influence which mechanism is dominant at any given temperature; at high temperature, where $1/k_B T$ is small, $(1/k_B T)^2 \gg (1/k_B T)^3$ and the EY mechanism is likely to be less important than the DP mechanism, while at low temperature, where $1/k_B T$ is large, $(1/k_B T)^2 \ll (1/k_B T)^3$ and the opposite is true.

### 5.3 Experiment

#### 5.3.1 Samples

All the samples studied in this work were provided by Qinetiq and BAE systems (either from in-house stock or purpose grown) with the aim of studying the effect of carrier type and concentration, and mobility on the electron spin lifetime in NGSs. Sample details can be found in Table 5.1.
### Table 5.1 Details of the InSb, HgCdTe, and InAs samples provided by Qinetiq and BAE systems for pump-probe spin relaxation measurements.

<table>
<thead>
<tr>
<th>Sample Code</th>
<th>Material</th>
<th>T (K)</th>
<th>Carrier Concentration (type, cm$^3$)</th>
<th>Mobility (cm$^2$V$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME1654</td>
<td>InSb</td>
<td>77</td>
<td>$n, 2.6 \times 10^{15}$</td>
<td>7870</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>$n, 1.5 \times 10^{16}$</td>
<td>68700</td>
</tr>
<tr>
<td>ME1655</td>
<td>InSb</td>
<td>77</td>
<td>$n, 2.8 \times 10^{15}$</td>
<td>29900</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>$n, 1.4 \times 10^{16}$</td>
<td>62700</td>
</tr>
<tr>
<td>ME1629</td>
<td>InSb</td>
<td>77</td>
<td>$n, 2.08 \times 10^{17}$</td>
<td>47700</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>$n, 2.26 \times 10^{17}$</td>
<td>41500</td>
</tr>
<tr>
<td>ME1722</td>
<td>InSb</td>
<td>77</td>
<td>$n, 2 \times 10^{17}$</td>
<td>45000</td>
</tr>
<tr>
<td>G256</td>
<td>Hg$<em>{0.78}$Cd$</em>{0.22}$Te</td>
<td>150</td>
<td>$n, 3 \times 10^{16}$</td>
<td>50000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>220</td>
<td></td>
<td>18000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td></td>
<td>7000</td>
</tr>
<tr>
<td>GVG763</td>
<td>Hg$<em>{0.7}$Cd$</em>{0.3}$Te</td>
<td>150</td>
<td>$p, 2 \times 10^{16}$</td>
<td>40000</td>
</tr>
<tr>
<td>GVG775</td>
<td>Hg$<em>{0.78}$Cd$</em>{0.22}$Te</td>
<td>150</td>
<td>$p, 1 \times 10^{17}$</td>
<td>20000</td>
</tr>
<tr>
<td>IC313</td>
<td>InAs</td>
<td>300</td>
<td>$n, 3.8 \times 10^{16}$</td>
<td>30000</td>
</tr>
<tr>
<td>IC311</td>
<td>InAs</td>
<td>300</td>
<td>$n, 1 \times 10^{17}$</td>
<td>25000</td>
</tr>
</tbody>
</table>

### 5.3.2 Time-Resolved Optical Pump-Probe Experiments

A time-resolved optical pump-probe measurement is so-called because a system is optically ‘pumped’ by an ultrafast laser pulse so as to induce some non-equilibrium state (in this case a spin-polarisation), which is subsequently measured some time later by a weaker time-delayed ‘probe’ ultrafast laser pulse. The delay introduced between the pump and probe pulse is usually controlled by means of delay arm, the length of which can be altered by means of a translatable retroreflector. A typical time-resolved pump-probe experimental arrangement, as used for spin relaxation measurements, is illustrated in Figure 5.5. The delay in that case is shown added to the pump path; it may be noted that advancing the pump beam by removing path length is equivalent to retarding the probe by adding path length. For spin relaxation measurements the timescales of interest are of the order of picoseconds, corresponding to a path difference between pump and probe of the order of tens of
millimetres. By scanning through a range of path differences, a spectrum of the probe transmission at the detector as a function of time (delay) can be constructed. In this way the temporal evolution (or dynamics) of carrier populations in a semiconductor can be studied.

Figure 5.5 A typical time-resolved MIR pump-probe arrangement for spin relaxation measurements. A linearly polarised laser pulse is split (unequally) into pump and probe pulses at the first beamsplitter. The probe beam is further split into an optional reference component for the purpose of real time normalisation of signal fluctuations due to laser instability. The probe and reference beams are then recombined with a long delay (compared to the duration of the process being studied; for carrier relaxation >10ns is usually adequate). A shorter delay (of the order of picoseconds or nanoseconds) delay is concurrently added to the pump beam, and both pump and probe (and reference) are focussed onto the sample. The pump beam is terminated at a beam stop, while the probe beam is detected.

Note generally a laser pulse is linearly polarised by nature. However, circularly polarised pulses are required to generate a spin-polarisation in a sample. These can be achieved by adding a quarter-wave plate circular polarisers to the pump and/or probe beams as required. As a circularly polarised probe beam is sensitive to the spin polarisation of a carrier population, such a beam can thus be used to probe spin relaxation in addition to carrier relaxation (note spin polarisations are discussed in more detail in the following section).

An optional reference arm can also be added to the set up as shown in Figure 5.5, its purpose being to minimise the effects of laser fluctuations on the measured signals. It is implemented by splitting the probe beam, and adding to one of the resultant paths
an additional length of the order of metres, chosen such that the corresponding time delay is sufficiently long that the system can be assumed to have returned to equilibrium by the time the reference arrives at the sample. The split beams are subsequently recombined temporally separated by the added delay. Any measured change in the reference pulse signal can thus be assumed to be due to laser instabilities, and the probe beam can be normalised appropriately on a pulse-by-pulse basis.

5.3.3 Spin Polarisation

In this work, optical interband excitation techniques have been employed at MIR wavelengths to create and probe polarised spins populations in NGSs. In a III-V semiconductor, for example, the angular momentum, $j$, of a conduction band Bloch state is 1/2, while for the heavy hole and light hole bands it is 3/2 and 1/2 respectively. Therefore, by irradiating the sample of interest with either right-handed ($\sigma^+$) or left-handed ($\sigma^-$) circularly polarised light of energy larger than the material bandgap carriers of preferentially one spin orientation can be excited.[15] Assuming for the sake of argument right-handed photons (i.e. $\sigma^+=+1$), angular momentum selection rules permit only two transitions; excitation of either spin-down electrons in the heavy hole band (i.e. $m_j=-3/2$) into spin-down conduction band states (i.e. $m_j=-1/2$), or spin-down electrons in the light hole band (i.e. $m_j=-1/2$) into spin up conduction band states (i.e. $m_j=+1/2$). Because the rates at which these two transitions can proceed is different, more electrons of one spin orientation will be excited into the valence band than the other. In fact, it can be shown that these absorptions occur in a 3:1 ratio, with the heavy hole transition more likely than the light hole transition.[8] Hence, (ideally) there will be three spin-down electrons in the conduction band for every one spin-up; a spin polarised population of electrons and holes has thus been created (the spins of the holes being opposite aligned to that of the electrons). For left handed photons (i.e. $\sigma^-=1$) the reverse of this polarisation is obtained. The degree of spin polarisation among electrons, $P_e$, is defined as follows:[8,16-18]

\[
P_e = \frac{n_\uparrow - n_\downarrow}{n_\uparrow + n_\downarrow} \quad (5.15)
\]
where \( n^\uparrow \) and \( n^\downarrow \) are the number of excited electrons in spin-up and spin-down states respectively. From equation 5.15 it can be deduced that the maximum possible degree of polarisation attainable is 0.5 (i.e. when \( n^\uparrow = 3 \) and \( n^\downarrow = 1 \)). The corresponding degree of polarisation for holes rapidly decays from this ideal value due to the strong spin-orbit coupling present in the valence band.

The re-randomisation of the spins in the valence band typically occurs on a timescale significantly faster than that of recombination, so only half of the recombining electrons with spin-up will give rise to \( \sigma^+ \) emission (and similarly for spin-down and \( \sigma^- \) emission). The net optical polarisation, \( P_{\text{opt}} \) (sometimes referred to as \( P \) in optical experiments), of the emitted light will therefore be only 0.25 (i.e. \( P_{\text{opt}} = P_e^2 \)).[8] However, that is an idealised result that assumes electron spins remain ideally polarised indefinitely. In reality, the electron spins will also randomise at (slower) rate \( 1/\tau_s \). If that rate is faster than, or comparable to, the recombination rate, \( \tau_R \), the degree of optical polarisation will be reduced by a factor of \( \tau_s(\tau_s + \tau_R)^{-1} \).

Based on the above principles, there are several variations on a common experimental theme designed to measure \( P_{\text{opt}} \). Those used in this work are the ‘Faraday rotation’ method and the ‘polarisation pump-probe’ technique.[16,17]

**5.3.3.1 Faraday Rotation Method**

Linearly polarised light may be considered a superposition of equal components of left- and right-handed circularly polarised light. When such a beam passes through a medium with a net magnetic moment, the phase velocity of those components are dependent on their orientation with respect to the magnetic field vector of the material.[19] The polarisation vectors of left- and right-circularly polarised light are separated by 180° by definition, so generally the angle they make with the field cannot be the same. As the polarisation angle of the linearly polarised superposed state is dependent on the relative phase of its circularly polarised constituents, the resulting difference in phase velocity within the medium must induce a continuous rotation of the (superposed beam’s) polarisation angle. The beam emergent from the
medium thus exhibits a net rotation\(^{17}\) (or phase shift), referred to as Faraday rotation for Michael Faraday who first identified the effect in 1845. The rotation angle, $\beta$, is related to the magnetic field strength and thickness of the medium by the following relation:[19,20]

$$\beta = \nu B d$$ (5.16)

where $d$ is the medium thickness, and $\nu$ is the Verdet constant. The latter is an empirical proportionality constant (units radT\(^{-1}\)m\(^{-1}\)) which varies with wavelength and temperature. Faraday rotation is illustrated schematically in Figure 5.6.[20]

![Figure 5.6 Schematic of Faraday Rotation in a magnetic medium. The polarisation of the linearly polarised beam is shown to rotate during its passage through the material, emerging at an angle $\beta$ with respect to the original angle of polarisation. The longer the path within the medium, the larger the observed $\beta$.][20]

In a pump-probe experiment, the spin-polarisation induced by the pump beam should have an associated net magnetic moment, such that when the probe beam passes through the sample some short time later it undergoes a Faraday rotation due to that field. It can be seen from equation 5.16 that this rotation should be proportional to the magnitude of that magnetic moment (via $B$) and, hence, the degree of electron polarisation in the sample. The rotation is therefore a direct gauge of $P_{\text{opt}}$.

Omitting the circular polariser in the probe beam path in Figure 5.5, by positioning an analysing linear polariser immediately before the detector the Faraday rotation of the linearly polarised probe beam can be measured. In other words, using this polariser to

\(^{17}\) In principle a beam could be rotated through $2\pi$ radians such that the polarisation angle of the exiting light does not appear rotated at all. However, the Faraday effect may still be expressed in terms of the phase shift of the circularly polarised components.
extinguish the probe signal when the pump is blocked, any signal subsequently
detected when the pump is unblocked must (in principle) be a result of the Faraday
rotation effect; $P_{\text{opt}}$ is thus directly measured. However, generally one can expect
some leakage through the polariser even in its fully extinguishing position. In that
case, a better option can be to perform separate measurements with the analysing
polariser orientated in the ±45° position (with respect to the probe polarisation)$^{18}$. The
difference between the signals in each polariser position then gives $P_{\text{opt}}$, with any
background signals arising due to the leakage of un-rotated components through the
analysing polariser cancelled out in the subtraction.

5.3.3.2 Polarisation Pump-Probe Technique

In the polarisation pump-probe technique, both the pump and probe must be circularly
polarised (as in Figure 5.5).[16] To obtain a measurement of $P_{\text{opt}}$, experiments are
performed with the beams’ circular polarisations equal (whether left-handed or right-
handed is not important), and with them opposite. Signals measured in the former
configuration are referred to as SCP (Same Circular Polarisation), and for the latter
OCP (Opposite Circular Polarisation). In the SCP case, (partial) bleaching of the
sample’s absorption by the pump results in greater transparency for the probe. The
resulting increase in probe signal intensity is, therefore, a measure of the number
electrons in the majority spin state. Conversely, in the OCP case the probe absorption
is increased, and is thus a probe of the number of minority spins in the sample. Hence,
if the pump is arranged with $\sigma^-$ polarisation, for SCP the change in probe (also $\sigma^-$)
absorption, $\Delta T_{\text{SCP}}$, is proportional to $3n_{\text{opt}}^{+} + n_{\text{opt}}^{-}$ (where $n_{\text{opt}}$ is the number of optically
generated electrons), while for OCP (probe $\sigma^+$) the change, $\Delta T_{\text{OCP}}$, is proportional to
$n_{\text{opt}}^{+} + 3n_{\text{opt}}^{-}$. The degree of optical polarisation $P_{\text{opt}}$ may, therefore, be written as
follows:[16]

$$P_{\text{opt}} = \frac{\Delta T_{\text{SCP}} - \Delta T_{\text{OCP}}}{\Delta T_{\text{SCP}} + \Delta T_{\text{OCP}}} \quad (5.17)$$

$^{18}$ This angle was chosen as the rate of change of a linear polarizer’s $\sin^2$ dependence as a function of
polarisation angle is peaked there. A given Faraday rotation will, therefore, produce a larger change in
signal at ±45°
where the numerator is the difference in probe transmission between SCP and OCP for left- and right-handed circularly polarised light, often referred to as the ‘circular dichroism’ (or CD) of the sample when a spin-polarised population is present. The denominator, on the other hand, is equivalent to the sum of all the carriers optically generated by the pump, $n_{\text{opt}}$, irrespective of polarisation, and represents the population decay due to electron-hole recombination. Any unpolarized electrons will contribute an absorption background which reduces $P_{\text{opt}}$ from its maximum value as follows:[16]

$$P_{\text{opt}} = 0.25 \left( 1 + \frac{n_b}{n_{\text{opt}}} \right) \quad (5.18)$$

where $n_b$ is the number of unpolarised electrons in the conduction band.[15]

The rate at which a given spin population imbalance ($\Delta n$) relaxes is given by the following rate equation:[18]

$$\frac{d\Delta n}{dt} = -\left[ \frac{1}{\tau_R} + \frac{2}{\tau_s} \right] \Delta n \quad (5.19)$$

where $(1/\tau_R + 2/\tau_s)$ is the population’s decay constant. In NGSs, Auger recombination usually dominates over radiative recombination (which will be discussed at greater length in chapter 7), so $1/\tau_R = Cn^2$, where $C$ is the Auger recombination coefficient. Rearranging equation 5.19 and integrating, the following expression for the spin population at time, $t$, is obtained:

$$\Delta n \propto \exp \left[ -\left( \frac{1}{\tau_R} + \frac{2}{\tau_s} \right) t \right] \quad (5.20)$$

If $\tau_R >> \tau_s$ the spin lifetime can be extracted from experimental data with minimal effort (i.e. $\tau_s$ dominates the decaying signal). The main advantage of the polarisation pump-probe technique is that the signals obtained are usually significantly larger for a given polarisation than in the Faraday rotation method. The polarisation pump-probe
technique is, therefore, considered the method of choice for spin relaxation measurements due to its larger SNR, although the Faraday rotation method may still be used to check the validity of results in case of any technique-related experimental artefacts.

However, in regimes of high doping the resulting Moss-Burstein shift necessitates pumping above a sample's band edge. As absorption higher in the bands is stronger due to the larger joint density of states, a larger concentration of photo-excited carriers are generated as a result, leading to a larger Auger recombination rate. This will add to the effect of the already large carrier density in highly doped samples, pushing the system further toward a regime where \( \tau_R \approx \tau_s \). Distinguishing the two exponential decays defined by equation 5.20 (due to spin relaxation and recombination) in experimental data can be extremely difficult in such a regime, if not impossible. Optical pump-probe studies of the effects of doping on spin lifetimes are, therefore, not without added complication.

5.3.4 Experimental Procedure

5.3.4.1 Arrangement I: FELIX

The InSb and Hg\(_{1-x}\)Cd\(_x\)Te samples in Table 5.1 were measured at FELIX (the details of which have been described previously in section 1.2.2). Experiments were performed in the 5-10\(\mu\)m wavelength region, using both Faraday rotation and polarisation pump-probe techniques. Lower limits for wavelength and pulse duration were imposed by FELIX itself to be \(\sim 4\mu\)m and \(\sim 1-5\)ps respectively. Probe (and reference) pulse detection was achieved using a nitrogen-cooled HgCdTe detector.

The basic arrangement in this case was as shown in Figure 5.5 (save for variations in polariser positioning), with the third reference beam implemented. The probe and reference were separated temporally by 20ns (\(\sim 3\)m additional path). A 25MHz (40\(\mu\)s period) radio-frequency (RF) signal generator was used to create an electrical sinusoidal waveform, which was subsequently electronically mixed with the detector signal such that probe pulses arrived at a peak of the sinusoidal signal wave and reference pulses at a trough. With the pump blocked, the probe and reference pulse
intensities should (ideally) be equal, so that the integral of the mixed waveform should be zero in that case. By adjusting the phase of the sinusoid that condition was indeed met. With the pump unblocked, any change in signal which produces a non-zero integral value must be induced by the pump (i.e. a spin-polarisation in the sample).

5.3.4.2 Arrangement II: Ti-Sapphire

The InAs samples in Table 5.1 were measured at the University of Surrey using the Coherent Ti:Sapphire ultrafast laser system (described in more detail in section 1.2.3). Broadly speaking, this arrangement was as depicted in Figure 5.5. It may be noted that the reference arm portion was not required, as the Ti:Sapphire laser system is rather more stable than FELIX.

The DFG output was set to 3.4\(\mu\)m, just above the accepted bandgap of InAs. Average power output was typically \(\sim 5\text{mW}\), with a focal spot on the sample of \(\sim 100\mu\)m. Signals were detected using a nitrogen-cooled InSb detector. The polarisation pump-probe method was employed exclusively on this occasion, although in one case an improvement to the arrangement was made by adding a ZnSe photo-elastic modulator into the probe path. This replaced the mechanical chopper in phase-sensitive detection, and enabled the polarisation of the probe to be modulated directly, yielding circular polarisations far closer to ideal than those possible via manual manipulation of variable quarter-wave plates. This modulator has since become a permanent fixture in pump-probe measurements utilising this laser system, although only the final measurement presented in this work was chronologically late enough to make use of it.

5.3.5 Results and Discussion

5.3.5.1 InSb

The first experiments in this study were on lightly n-type doped \((n\sim 1.4-1.6\times10^{16}\text{cm}^{-3}\) at 300K) InSb samples ME1654 and ME1655. Using the polarisation pump-probe technique, a short spin lifetime of only \(\tau_s\sim 16\text{ps}\) has been determined for both samples.
at 300K. That lifetime is also found to be quite temperature insensitive, with the same value measured between 77K and 300K. The result on ME1655 at 150K is shown in Figure 5.7. The initial degree of polarisation in that case was determined to be $P \approx 0.2$, which is close to the theoretical limit (i.e. 0.25), lending some confidence to the measured lifetime. Note that at these long wavelengths the quarter-wave plates used are expected to produce imperfect circular polarisations, so optical polarisations smaller than the theoretical limit were anticipated.

**Figure 5.7** Pump-probe transmission change as a function of probe delay for lightly doped ($n=2 \times 10^{15} \text{cm}^{-3}$) InSb sample ME1655 at 150K, as measured using the polarisation pump-probe technique. Shown inset is the extracted optical polarisation as a function of probe delay. A mono-exponential fit (solid curve) to the resulting decay yields a measured lifetime of $\tau_s = 16 \text{ps}$.

The same procedure was repeated for more heavily $n$-type doped ($n=2 \times 10^{17} \text{cm}^{-3}$) sample ME1629. This yielded a substantially longer spin lifetime $\tau_s \approx 300 \text{ps}$ at 300K. However, that measurement was complicated by the presence of a large value for $n_b$, which reduced the size of the optically measured polarisation by a factor of $\approx 5$, in accordance with equation 5.18. To check that this result was not an artefact of the polarisation pump-probe technique, it was repeated using the Faraday rotation method. The result obtained is shown in Figure 5.8, and is found to be consistent with the previous measurement. A value for the optically generated carrier concentration in ME1629 was estimated to be $n_{\text{opt}} \approx 5 \times 10^{16} \text{cm}^{-3}$, larger than that for ME1655. This increase in $n_{\text{opt}}$ was anticipated, as the higher doping level of ME1629 results in a considerable Moss-Burstein shift. Therefore, it was necessary to pump above the band edge (where the joint density of states is larger), resulting in a correspondingly
stronger absorption of the FELIX excitation. Assuming Auger recombination to be the dominant carrier recombination mechanism in these (narrow gap) materials, the larger value of \( n_{\text{opt}} \) in ME1629 should result in a considerably shorter carrier lifetime, \( \tau_R \) (see section 7.2.1 for a more detailed discussion of Auger recombination). Indeed, inspection of Figures 5.7 and 5.8 shows that this does appear to be the case, with a more pronounced background decay observed in Figure 5.8. The carrier recombination rate is, therefore, given by \( 1/\tau_R \approx C(n_b+n_{\text{opt}})^2 \), where for these materials \( C \approx 2 \times 10^{-26} \text{cm}^6 \text{s}^{-1} \).\{22\} Values for \( \tau_R \) of 20ns and 800ps for ME1655 and ME1629 respectively are thus obtained, in good agreement with the average experimental decay time, \( (\Delta T_{\text{SCP}}+\Delta T_{\text{OCP}})/2 \), in each case. Temperature dependent measurements of the spin lifetimes in samples ME1629 and ME1722 (made in the range 77-300K) once again yielded an approximately constant value (i.e. \( \tau_s \approx 300 \text{ps} \) at all sampled temperatures).

**Figure 5.8** Pump-probe transmission change for heavily doped \( (n=2 \times 10^{17} \text{cm}^{-3}) \) sample ME1629 at 300K, obtained using the Faraday rotation method. The analysing polariser was set at ±45° with respect to the linear polarisation of the probe. Inset is shown the decay of the Faraday rotation angle as a function of time delay. A mono-exponential fit (solid curve) to the decay yields a measured spin lifetime \( \tau_s = 300 \text{ps} \) at 300K.

In the degenerate regime, the temperature dependence of \( \tau_s \) is primarily determined by the dependence of \( \tau_p \). For ME1629 and ME1722, the observed insensitivity of \( \tau_s \) to changes in temperature can thus be explained by their similarly temperature insensitive mobility (see Table 5.1). This is attributed to the opposite trends of ionised impurity scattering and lattice scattering with temperature; the former is proportional
to $T^{3/2}$, while the latter goes as $T^{-3/2}$. By coincidence it appears the constants of proportionality for each process are approximately equal in ME1629 and ME1722, so that changes in their contributions to the total sample mobility approximately cancel each other out.

However, for non-degenerate sample ME1655 equations 5.12 (DP mechanism) and 5.14 (EY mechanism) predict a lifetime with a strong temperature dependence (via their $(k_B T)^3$ and $(k_B T)^2$ terms respectively). Therefore, the insensitivity of the spin lifetime in that sample is not so easily explained. It is believed that in InSb, the narrowness of the energy gap means that even at low $n$-type doping levels the Fermi energy lies relatively close to the conduction band edge (e.g. at room temperature $<90$meV). Therefore, unlike in wide gap semiconductors, where there is a more clearly defined non-degenerate regime (i.e. in intrinsic material the Fermi level is typically several hundred meV from the conduction band edge), in InSb low doping levels correspond to a regime intermediate between degenerate and non-degenerate. Hence, the lifetime in that case should be more accurately defined by a rather complicated amalgamation of equations 5.11 and 5.12 (for the DP mechanism), and equations 5.13 and 5.14 (for the EY mechanism). The strong temperature dependence of equations 5.12 and 5.14 (non-degenerate) may, therefore, be diluted somewhat, possibly leading to a temperature dependence more like that predicted by equations 5.11 and 5.13 (degenerate).

For ME1655 (the non-degenerate case) at 300K, reliable literature values for some of the parameters in equations 5.12 (DP) and 5.14 (EY) are available; $E_g(T=300K)=0.17eV$, $\Delta_{so}=0.81eV$, and $m_c/m_0=0.14$ at the conduction band edge. The momentum scattering time, $\tau_p$, can be estimated from the mobility, $\mu$, using the following well-known expression:[23]

$$\mu = \frac{e \tau_p}{m_c} \quad (5.21)$$

where $\mu$ has been measured at 300K and is given in Table 5.1 ($\mu=62700cm^2/Vs$), and $e$ is the electronic charge. Rearranging equation 5.21 for $\tau_p$, a value of 0.5ps is
obtained. Assuming the dominance of lattice scattering processes in this case, the
dimensionless scattering constants in equations 5.12 (DP) and 5.14 (EY) are \( \Gamma' = 0.8 \)
and \( \Lambda' = 2 \) respectively.[15] These values yield predicted spin lifetimes for ME1655 for
the DP and EY mechanism of 10.8ps and 24.0ps respectively. The experimentally
determined lifetime of 16ps lies close to the middle of the range defined by those two
values, possibly indicating comparable contributions from both mechanisms to spin
relaxation.

A similar calculation for degenerately-doped sample ME1629 at 300K
(\( \mu = 41500 \text{cm}^2\text{V}^{-1}\text{s}^{-1} \)) is rather more troublesome, as non-parabolicity of the InSb
conduction band is expected to be important in that case. This makes accurate
determination of \( E_F \) and the relevant value of \( m_c/m_0 \) difficult. In particular, because \( E_F \)
appears cubed in equation 5.11, the spin lifetime predicted by the DP is extremely
sensitive to the value used. For example, a reduction of \( E_F \) (believed to be \( \sim 0.03 \text{eV} \))
by \( \sim 30\% \) is found to result in a \( \sim 300\% \) increase in the DP spin lifetime. Although the
dependence of the EY mechanism on \( E_F \) is weaker (proportional to the square), a
\( \sim 200\% \) increase in the EY lifetime is still observed for a similar adjustment. Any such
calculations are, therefore, unreliable. Nevertheless, using the estimations \( E_F = 0.03 \text{eV} \)
and \( m_c/m_0 = 0.027 \) (based on previous work by co-workers) [25], and values for the
scattering constants (assuming dominance of ionised impurity scattering in this
degenerate case) of \( \Gamma = 16/315 \) (DP, equation 5.11) and \( \Lambda = 32/27 \) (EY, equation 5.13),
spin lifetimes of \( \sim 23.0\text{ps} \) and \( \sim 38.5\text{ps} \) are obtained for the DP and EY mechanisms
respectively.[15] Although caution must be exercised before reading too much into
that result, and the values are considerably smaller than the experimentally
determined lifetime of \( \tau_\text{s} \sim 300\text{ps} \), it appears clear that both the experimental and
calculated lifetimes are longer for the degenerate sample than the intrinsic sample,
indicating at least a similar trend.

It should be noted that the pre-factors in both DP and EY theoretical expressions, are
considerably different in the degenerate and non-degenerate regimes. It is possible
that the non-degenerate pre-factors (assuming dominance of lattice scattering
processes) are accurate (hence the good agreement with experiment), while the
degenerate pre-factors (assuming dominance of ionized impurity scattering) are in
error, thus explaining the order of magnitude discrepancy between experiment and theory. More samples would need to be measured demonstrating repeatability of the ME1629 result to verify this claim, however.

In any case, these measurements indicate a strong suppression of spin relaxation in degenerately doped InSb, although the mechanism for this suppression is not currently understood. Additionally, the results presented here include the first determinations of spin lifetimes in InSb at room temperature, and are found to be of the same order of magnitude as earlier results measured at low temperature.[15]

5.3.5.2 Hg$_{1-x}$Cd$_x$Te

Although the bandgap of Hg$_{1-x}$Cd$_x$Te for 0.2≤x≤0.3 is only slightly smaller than that of InSb, it is not possible to find a temperature below 300K at which the parameter is the same in both, as the temperature variation of II-VI alloys such as HgCdTe is the opposite of that exhibited by III-V semiconductors.

The first Hg$_{1-x}$Cd$_x$Te measurements were on low p-type doped (p=2×10$^{16}$cm$^{-3}$) sample GVG763 (x=0.3), which produced a short lifetime of $\tau_s\sim$20ps. This result appears consistent with that obtained for lightly n-type doped InSb (i.e. ME1654 and ME1655). A similar measurement on the more heavily p-type doped (p=1×10$^{17}$cm$^{-3}$) sample GVG775 (x=0.22) yielded an even shorter lifetime of just $\tau_s\sim$5ps. It may be noted that this value was probably resolution limited due to the FELIX pulse duration (of the order of picoseconds) though.

On moving to the moderately n-type doped (n=3×10$^{16}$cm$^{-3}$) sample G256 (x=0.22), a much longer lifetime $\tau_s=96$ps (at 220K) was found, broadly fitting the trend of lengthening spin lifetime with increasing n-type doping levels observed in InSb. This polarisation pump-probe measurement is shown in Figure 5.9.
Figure 5.9 Pump-probe transmission change as a function of probe delay for moderately doped \((n=3 \times 10^{16} \text{cm}^{-3})\) Hg\(_{0.78}\)Cd\(_{0.22}\)Te sample G256 at 220K, as measured using the polarisation pump-probe technique. Shown on the inset plot is the extracted optical polarisation as a function of probe delay. The mono-exponential fit (solid curve) to the resulting decay yields a measured lifetime of \(\tau_s=96\text{ps}\).

A temperature dependent measurement of the spin relaxation lifetime in this sample was also performed (at 150K, 220K and 300K), the results of which are shown in Figure 5.10. In this instance (i.e. non-degenerate statistics) equation 5.12 (DP) and equation 5.14 (EY) are expected to apply. Also, in this non-degenerate limit \(\tau_p\) is expected to be dominated by lattice scattering processes, so its temperature variation is anticipated to go as \(T^{3/2}\).\(^{[2,25]}\) Incorporating this dependence into equation 5.14 for the EY mechanism, one obtains \(T^{-7/2}\) dependence of \(\tau_s\), corresponding to a factor of \(~11.3\) decrease between 150K and 300K. A similar substitution in equation 5.12 for the DP mechanism yields a \(T^{3/2}\) dependence of \(\tau_s\), giving a reduction of only a factor of \(~3\) over the same temperature range. The factor \(~14.8\) decrease of the spin lifetime in G256 over that temperature range is, therefore, more consistent with the prediction of the EY relaxation mechanism. Table 5.2 lists the theoretically predicted and experimental changes in spin lifetimes for three different temperature steps, based on the three temperatures measured. There it can be seen that the prediction of the EY mechanism is consistently in better agreement in each case. Therefore, it is concluded that the EY mechanism dominates in moderately doped Hg\(_{0.78}\)Cd\(_{0.22}\)Te.
<table>
<thead>
<tr>
<th>$T$ Step</th>
<th>Exp. Increase</th>
<th>DP Predicted Increase ($\propto T^{3/2}$)</th>
<th>EY Predicted Increase ($\propto T^{-7/2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>150-220</td>
<td>3.7</td>
<td>1.8</td>
<td>3.8</td>
</tr>
<tr>
<td>220-300</td>
<td>4</td>
<td>1.6</td>
<td>3</td>
</tr>
<tr>
<td>150-300</td>
<td>14.8</td>
<td>2.8</td>
<td>11.3</td>
</tr>
</tbody>
</table>

Table 5.2 Comparison of experimentally observed and theoretically predicted increases in spin lifetime in non-degenerately $n$-type doped ($n\approx3\times10^{16}$cm$^{-3}$) Hg$_{0.76}$Cd$_{0.22}$Te sample G256 for three different temperature steps. The prediction of the EY model (equation 5.14) is consistently in better agreement with the experimentally measured values, indicating dominance of the EY mechanism in this case.

Figure 5.10 Temperature dependence study of spin-polarisation decay in non-degenerate ($n\approx3\times10^{16}$cm$^{-3}$ at 150K) Hg$_{0.76}$Cd$_{0.22}$Te sample G256, as measured using the polarisation pump-probe technique. The decay of optical polarisation is plotted for three temperatures; 150K, 220K, and 300K. The lifetimes extracted from the data at each temperature by means of fitting mono-exponential functions (solid curve) are found to be in good agreement with theoretical predictions (i.e. equation 5.14).

5.3.5.3 InAs

The first InAs sample to be measured in this study was IC313, a lightly $n$-type doped sample ($n\approx3.8\times10^{16}$cm$^{-3}$). The results of the polarisation pump-probe experiment on that sample are shown in Figure 5.11, and appear to be in good agreement with those obtained by Boggess et al on a similarly doped sample ($n\approx2\times10^{16}$cm$^{-3}$). The initial degree of polarisation was found to be only $P\approx0.09$, well below the ideal value of 0.25. This discrepancy is once again attributed to imperfect circular polarisation of the pump and probe beams, both due to the circular polarisers themselves, and reflections.
from mirrored optics positioned between the polarisers and the sample. A spin lifetime of $\tau_s=20\pm1$ps has been extracted from the decay, again in good agreement with the value of 19ps determined by Boggess et al.[16]

![Figure 5.11](image)

**Figure 5.11** Pump-probe transmission change as a function of probe delay for lightly doped ($n=3.8\times10^{16}$cm$^{-3}$) InAs sample IC313 at 300K, as measured using the polarisation pump-probe technique. Shown on the inset plot is the extracted optical polarisation as a function of probe delay. The resulting mono-exponential decay (solid curve) yields a lifetime of $\tau_s=20$ps.

Following that measurement, the effect of near degenerate $n$-type doping levels was investigated using sample IC311 ($n\approx1\times10^{17}$cm$^{-3}$). At this point the arrangement was improved with the addition of a ZnSe photo-elastic modulator to the probe path (in place of a quarter-wave plate), as described in section 5.3.4.2. Further, it was ensured on this occasion that the modulator was the last optical component encountered by the beam before reaching the sample. When arranged in this manner, the signal measured by the lock-in amplifier is equivalent to the difference in transmission for SCP and OCP polarisations, or circular dichroism (i.e. $\Delta T_{CD}=\Delta T_{OCP}-\Delta T_{SCP}$). The population decay due to electron-hole recombination may be similarly measured with the photo-elastic modulator off, and (mechanically chopped) pump and probe linearly polarised (i.e. $\Delta T_{LP}=T_{SCP}+T_{OCP}$). The degree of optical polarisation may thus be computed as $\Delta T_{CD}/\Delta T_{LP}$, the exponential decay of which yields a spin lifetime of $\sim24\pm2$ps. This result is shown in Figure 5.12. Within error limits, the spin lifetimes in both InAs samples at 300K appear indistinguishable based on these measurements.
Figure 5.12 Decay of optical polarisation in near-degenerate n-type \((n \approx 1 \times 10^{17} \text{cm}^{-3})\) InAs sample IC311, as measured using a ZnSe photo-elastic modulator in an otherwise conventional polarisation pump-probe arrangement. Shown inset are the difference between SCP and OCP transmission (referred to as circular dichroism, or CD), as obtained direct from the lock-in output (solid dots). Also plotted is the population decay due to electron-hole recombination (open dots), measured using linearly polarised (LP) pump and probe beams (equivalent to the sum of SCP and OCP). The degree of optical polarisation is the ratio CD/LP, with a fit to the resulting mono-exponential decay (solid line) yielding a spin lifetime of \(\tau_s \approx 24\text{ps}\).

However, a more recent measurement by colleagues on sample IC313 using the photo-elastic modulator-based technique has yielded a more precise lifetime of \(\tau_s \approx 16 \pm 2\text{ps}\).\[26\] Any suppression of spin relaxation in near-degenerate InAs is, therefore, considerably smaller than that observed in the earlier measurements on degenerate InSb. The reason for this discrepancy is not clear at this time, although attempts to investigate this finding further are already underway.

5.3.6 Conclusions and Future Work

Spin lifetimes have been measured for different doping levels in InSb and InAs. In InSb a strong suppression of spin relaxation at 300K (~30 times) has been observed in moving from non-degenerate to degenerate \(n\)-type doping levels. For the non-degenerate case, the measured lifetime at 300K is consistent with theory, and suggests comparable contributions of DP and EY to spin relaxation. The measured lifetimes weak dependence on temperature is not consistent with theory, however. This may indicate that InSb exists, even at low doping levels, in a regime intermediate between those described by the non-degenerate and degenerate formulations of the DP and EY
models. Additionally, the observed suppression in the degenerately doped sample is larger than that predicted by the theoretical expressions, highlighting a possible error in their scattering mechanism dependent pre-factor. More measurements on InSb with high doping levels are required to verify if this is indeed the case. The mechanism for the observed suppression is, therefore, unclear at this time. A far smaller suppression (<2 times) has been observed in near-degenerately n-type doped InAs at 300K.

Temperature dependent measurements of the spin lifetime in moderately n-type doped HgCdTe have yielded changes in the spin lifetime consistent with dominance of the EY relaxation mechanism. This finding is independent of the pre-factors in the theoretical expressions, so this conclusion is unaffected by any doubt regarding their validity.

Measurements on more samples with a variety of doping levels are ideally required, both to demonstrate repeatability of these results, and obtain a more complete understanding of the doping dependence of the spin lifetimes.

5.4 References


6 First Observation of the Spin Galvanic Effect in Zero Applied Magnetic Field

6.1 Introduction

Generally speaking, a galvanic effect is any mechanism which induces a current within a material. For example, shining light on certain semiconductor structures can result in a ‘photo’-galvanic effect. Occurring in semiconductor structures and materials possessing very particular symmetry properties, the so-called spin galvanic effect (SGE) is driven by spin selective scattering of a polarised electron gas.[1] The spin polarisation of this electron gas may be achieved by any available means (e.g. electrical injection, optical excitation, etc.). Ultimately it is hoped the SGE can be implemented in a semiconductor-based spintronic device by means of the reverse process, whereby an injected current generates a spin polarisation. Furthermore, evaluation of the effect can provide a useful diagnostic tool for evaluating the relative contributions of certain bulk and structural spin-splitting contributions, which may prove helpful for spintronic device engineering in the future.

At this stage of development, a common method of generating the spin polarisation necessary for the SGE (as a means of studying the effect) is optical excitation by circularly polarised photons. However, under conditions conducive to the generation of an SGE current, this method can simultaneously give rise to a second, less technologically interesting galvanic effect. Known as the circular photo-galvanic effect (CPGE), this arises from a spin-dependent asymmetry in the photo-excitation
The ability to differentiate between the contributions made by each effect to any photo-induced current is critical if one is to unambiguously identify and study the SGE. Previous demonstrations of the SGE have employed strong magnetic fields to separate the SGE and CPGE currents. The aim of this work is to use the tunability of FELIX (see section 1.2.2) to isolate the SGE current in [001] grown GaAs quantum well structures, without the application of an applied magnetic field.

### 6.2 Theory

#### 6.2.1 Linear Spin-Splitting in III-V Quantum Well Structures

Cubic and (to a lesser extent) linear spin-splittings in bulk III-V semiconductors were discussed previously in section 5.2.1. There, the role of bulk crystal symmetry in spin-orbit coupling interactions was described, with the lack of inversion asymmetry in bulk III-V semiconductors cited as key to the presence of such odd power $k$-dependent terms in the crystal Hamiltonian. This form of inversion asymmetry is usually referred to as bulk inversion asymmetry (BIA).

Additional inversion asymmetry may be introduced by growing lower dimensional heterostructures such as a quantum well (QW). For example, asymmetric growth (e.g. non-uniform width, different barrier heights on either side, etc.), non-uniform doping, and applied electric fields may all be used to ensure a grown QW structure lacks inversion symmetry. This form of inversion asymmetry is usually referred to as structural inversion asymmetry (SIA). It may also be noted that the overall symmetry of a QW is dependent on the crystal plane in which it is grown.

A further source of inversion asymmetry may be provided if the atomic layers at each of the QW barrier interfaces are different. For example, a III-V semiconductor grown in the (100) crystal plane alternates atomic species with each growth layer. Therefore, by carefully controlling the QW growth, the first atomic layer can be engineered to be group III, while the final layer is made group V. This form of inversion symmetry is referred to as Interface Inversion Asymmetry (IIA), though IIA-related terms in the Hamiltonian of an electron within the well are found to be indistinguishable from those arising from BIA. Such terms are referred to collectively as Dresselhaus
terms, and were defined previously for a bulk III-V semiconductor in section 5.2.1. SIA contributes so-called Rashba terms to the electron Hamiltonian. It may be noted that IIA and SIA can be deliberately (and controllably) engineered during crystal growth, unlike BIA.

The GaAs QWs measured in this work are grown in the (001) plane, corresponding to the C2v symmetry point group.[10] Recall in section 5.2.1 the effective magnetic field, $B_{\text{eff}}(k)$, representing a spin-orbit interaction was described in terms of precession vector, $\Omega(k)$ (equation 6.7). An expression for $\Omega(k)$ due to SIA has been derived by Bychkov and Rashba:[8]

$$\Omega(k) = 2\alpha(k \times n) \quad (6.1)$$

where $\alpha$ is the Rashba term, $n$ is a vector normal to the plane of the QW, and wavevector $k=(k_x, k_y)$ where $k_x$ and $k_y$ define orthogonal directions in the plane of the QW. In this instance, $x$ and $y$ are directed along the [100] and [010] directions respectively. The magnitude of the spin-orbit interaction is contained within $\alpha$. The spin-orbit term in the electron Hamiltonian, $\mathcal{H}_{\text{so}}$ (where $\mathcal{H}=\hbar^2 k^2/2m+\mathcal{H}_{\text{so}}$), is related to $\Omega(k)$ via the following relation:[10]

$$\mathcal{H}_{\text{so}}(k)=\frac{\hbar}{2}\sigma \cdot \Omega(k) \quad (6.2)$$

where $\sigma$ is the vector of the Pauli matrices. For (001) grown QWs $n||\langle001\rangle$, so combining equations 6.1 and 6.2 one obtains the following expression for the SIA contribution to $\mathcal{H}_{\text{so}}$:

$$\mathcal{H}_{\text{so}} = \alpha(\sigma_y k_x - \sigma_x k_y) \quad (6.3)$$

where the additional factor of $\hbar$ in equation 6.2 has been incorporated into $\alpha$. A similar expression can be found for the BIA/IIA contribution to $\mathcal{H}$, only the equivalent of equation 6.1 in that instance contains the scalar product rather than the cross product. The spin-orbit interaction term may then be written in full as the sum of
SIA and BIA/IIA contributions as follows:[4]

\[ \mathcal{H}_{so} = \alpha \left( \sigma_y k_y - \sigma_x k_x \right) + \beta \left( \sigma_y k_x - \sigma_x k_y \right) \]  

(6.4)

where \( \beta \) is the Dresselhaus term. In sufficiently narrow quantum wells, the expectation value of the component of an electron's wavevector perpendicular to the well plane, \( k_n^2 = \langle (k \cdot n)^2 \rangle \), is much larger than the in-plane components (i.e. \( k_n^2 \gg k_x, k_y \)).[10] As a result, cubic terms involving the in-plane directions (e.g. \( k_x k_y^2, k_y k_x^2, \ldots \)) can be neglected and one is left with terms linear in \( |k| \) multiplied by the constant \( k_n^2 \), the latter being incorporated into the Rashba term for a given structure (as in equation 6.4).

The effective magnetic fields are easily obtained from equation 6.4 (see section 5.2.1 where this was demonstrated in the case of cubic spin-splittings). In the limits of pure Rashba (i.e. \( \alpha \neq 0, \beta = 0 \)) and pure Dresselhaus splitting (i.e. \( \alpha = 0, \beta \neq 0 \)) these are found to be \( B_{\text{eff}}(k) = \alpha (k_y, -k_x) \) and \( B_{\text{eff}}(k) = \beta (k_x, -k_y) \) respectively. The resulting distribution of spin-eigenstates in each case is illustrated in Figure 6.1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6_1.png}
\caption{Schematic showing the distribution of spin-eigenstates (or orientations of \( B_{\text{eff}}(k) \)) resulting from pure Rashba terms (SIA) and pure Dresselhaus (BIA and IIA) terms in \( \mathcal{H}_{so} \). All spins are aligned in-plane in both cases, though the two splitting mechanisms do result in different in-plane spin alignments.}
\end{figure}

In the presence of only Rashba or Dresselhaus terms (i.e. \( \alpha = 0, \beta \neq 0 \) or \( \alpha \neq 0, \beta = 0 \)), the electron Hamiltonian, \( \mathcal{H} \), is given by:
\[
\frac{\hbar^2 k^2}{2m^*} \pm \sqrt{2|\alpha|k} \quad \text{for pure SIA} \quad (6.5)
\]
\[
\frac{\hbar^2 k^2}{2m^*} \pm \sqrt{2|\beta|k} \quad \text{for pure BIA/IIA} \quad (6.6)
\]

where the spin-orbit terms in each case follows from equation 6.4. Hence, aside from the magnitude of the spin-splitting (as defined by \(\alpha\) and \(\beta\) respectively), the \(E-k\) dispersions resulting from purely SIA or BIA induced spin-orbit coupling are identical, as illustrated in Figure 6.2. The only difference between the two bandstructures is the orientation of the resulting spin-eigenstates. By combining the information presented in Figure 6.1 and Figure 6.2, it is hopefully apparent that the additional \(k\)-linear spin-orbit term in equation 6.5 and equation 6.6 lead to a linear shift of the spin sub-bands about \(|k|=0\). Furthermore, when only one of BIA or SIA is present, that shift is isotropic in all in-plane (with respect to the QW) directions.

\textbf{Figure 6.2} Schematic illustrating a typical \(E-k\) dispersion resulting from pure Rashba terms (SIA) in \(\mathcal{H}_0\). The dispersion is shown in 2D on the RHS. The green and red surfaces represent the positive and negative solutions for \(k_{\parallel}\). The dispersion is seen to be isotropic in all in-plane directions. The dispersion in 1D, specifically for \(k_{|[100]}\), is shown on the RHS, although all other directions are identical. Coloured arrows indicate the orientation of spin-eigenstates along [100] in the spin-split sub-bands (see also Figure 6.1). The \(E-k\) dispersion resulting from pure Dresselhaus terms is identical.

In the more general case, where both Rashba and Dresselhaus terms are non-zero, the effective magnetic field is found to be \(\mathbf{B}_{\text{eff}}(\mathbf{k})=(\alpha k_y+\beta k_z, -\alpha k_z-\beta k_y)\). The corresponding
eigenstate distribution is shown in Figure 6.3, where it is also clear that the [110] and [1-10] directions are now non-equivalent.

![Figure 6.3 Schematic showing the distribution of spin-eigenstates (or orientation of $B_{\text{eff}}(k)$) resulting from the presence of both Rashba (SIA) and Dresselhaus (BIA and IIA) terms in $\mathcal{H}_{so}$. All spins are aligned in-plane. Their orientations are now a weighted mixture of those arising from only one or other splitting. Along [110] and [1-10] the spin-eigenstates align perpendicular to $k$, irrespective of the (non-zero) magnitudes of $\alpha$ and $\beta$.](image)

The Hamiltonian along the [110] and [1-10] directions is given by:

\[
\frac{\hbar^2 k^2}{2m^*} \pm \sqrt{2} (\alpha - \beta) |k| \quad \text{for } k||[110] \quad (6.7)
\]

\[
\frac{\hbar^2 k^2}{2m^*} \pm \sqrt{2} (\alpha + \beta) |k| \quad \text{for } k||[1-10] \quad (6.8)
\]

The magnitude of the splitting in the [110] and [1-10] each direction is, therefore, different; for $k||[110]$ it is proportional to $(\alpha+\beta)$, while for $k||[1-10]$ it is proportional to $(\alpha-\beta)$. The equivalent expressions for the [100] and [010] directions are identical to one another, and read as follows:

\[
\frac{\hbar^2 k^2}{2m^*} \pm \sqrt{2(\alpha^2 + \beta^2)} |k| \quad \text{for } k||(100) \text{ or } k||(010) \quad (6.9)
\]

The resulting $E-k$ dispersion for non-zero Rashba and Dresselhaus terms is illustrated
in Figure 6.4. The non-equivalence of the [110] and [1-10] directions when $\alpha \neq \beta \neq 0$ is of paramount importance to this work, as it provides a means of isolating the SGE and CPGE currents. How this is done will be made clearer by the following discussions of each effect.

Consider a hypothetical sample (of the above structure), cleaved quadratic in shape, with the $x$ and $y$ directions arranged to lie along [1-10] and [110] respectively (i.e. the axes of Figures 6.3 and 6.4 are rotated 45° clockwise). Contacts allowing the sample
to be connected to an external circuit (for current measuring purposes) are positioned in the middle of each of the four sample sides. This sample geometry is illustrated in Figure 6.5. Illuminating the sample at an oblique incidence with circularly polarised light (assume $\sigma^+$ for the sake of argument), the photon energy of which is suitable for intra-conduction band transitions between QW confined states (or sub-bands) $e_1$ (the ground state) and $e_2$ (the first excited state), a spin polarisation, $S$, may be generated co-parallel with the photon's polarisation vector, $\hat{e}$. For example, if the in-plane component of $\hat{e}$ is $\hat{e}_x || [1-10]$, the in-plane component of the resultant spin polarisation is $S_x || [1-10]$ also, as shown in Figure 6.5. The energy separation between the confined states in the QW is $E_{21}$.

In (001) grown QWs, selection rules require that the absorption of a circularly polarised photon by an electron in $e_1$ must be 'spin conserving'. That is, the spin of the electron itself is conserved.[15] Furthermore, for photons with an oblique angle of incidence, intraband absorption processes are found to be spin-dependent, with the probability of such an absorption being greater if $S_x$ is co-parallel with $\hat{e}_x$ than if it is aligned anti-parallel.[16]

Figure 6.5 Schematic illustrating the sample geometry for a hypothetical experiment demonstrating the CPGE. The sample is quadratic in shape, with contacts on each side to measure any CPGE induced current. Circularly polarised photons are incident at an oblique angle with in-plane polarisation projection $\hat{e}_x$. A net spin-polarisation $S_x$ is thus generated, which (due to linear spin-splitting in the QW bandstructure) leads to a current in the $y$-direction, $I_{CPGE}$.  

167
In the bandstructure of Figure 6.3 (i.e. \( \alpha \neq \beta \neq 0 \)), only electrons whose wavevectors are parallel to \( k_y \) have spin-eigenstates orientated in the \( x \)-direction. For electrons moving along all other directions precession of their spins will tend to average their net polarisations to zero. Considering then \( k || k_y \) only, for a given exciting photon energy, \( \hbar \omega < E_{21} \), there are only two possible transitions between \( e_1 \) and \( e_2 \) which both conserve energy and obey the relevant selection rules; those are found at \( \pm k_y \) respectively. This is illustrated on the LHS of Figure 6.6. For \( \sigma^+ \) excitation it is spin-up electrons which are aligned co-parallel with \( \hat{e}_x \), so the transition between the spin-up sub-bands dominates (as represented by a thicker arrow in Figure 6.6). The spin-up \( e_1 \) sub-band is thus emptied of electrons faster than the spin-down \( e_1 \) sub-band, with each emptied state leaving behind it an unexcited electron on the opposite side of its sub-band. As the group velocity of an electronic state is proportional to \( d\omega/dk_y \), the generated imbalance in each \( e_1 \) sub-band's population results in a current. The currents generated by each sub-band are unequal and oppose one another, with the larger current, \( I_{up} \), originating from the spin-up sub-band (due to the greater population imbalance), and the smaller current, \( I_{down} \), due to the spin-down sub-band. A net current \( I_{CPGE} = I_{up} + I_{down} \) (where in this instance \( I_{down} \) negative) thus flows along the [110] direction (i.e. perpendicular to \( \hat{e}_x \)).

Note that a similar population imbalance is also generated in the \( e_2 \) sub-bands after the initial excitation. However, the energy of the states in \( e_2 \) is well above the threshold energy for optical phonons in GaAs, so electrons in those sub-bands rapidly empty into \( e_1 \), neutralising the imbalance.\[{17}\) Hence, any current generated in \( e_2 \) is rapidly dissipated, allowing its contribution to \( I_{CPGE} \) to be neglected.
First Observation of the Spin-Galvanic Effect in Zero-Applied Magnetic Field

\[ I_{CPGE} = I_{up} + I_{down} \]

\[ \sigma^+ \quad h_0 \omega = E_{21} \]

\[ e_1, \quad j_x = +1/2 \]

\[ e_2, \quad j_x = -1/2 \]

\[ e_2, \quad j_x = +1/2 \]

\[ e_1, \quad j_x = -1/2 \]

\[ E_{21} \]

\[ \sigma^+ \quad h_0 \omega > E_{21} \]

\[ e_1, \quad j_x = +1/2 \]

\[ e_2, \quad j_x = -1/2 \]

\[ I_{down} \quad e_2 \]

\[ I_{up} \quad e_1 \]

Figure 6.6 Schematic illustrating the microscopic origin of the CPGE effect in an (001) grown QW for \( \sigma^+ \) incident light with in-plane polarisation \( \hat{e}_x \). A spin-polarisation and population imbalance is generated by a spin-dependent asymmetric excitation of electrons between spin sub-bands \( e_1 \) (ground state) and \( e_2 \) (first excited state), leading to currents \( I_{up} \) and \( I_{down} \) proportional to \( \frac{\partial \alpha}{\partial k_y} \) (in \( e_1 \)). The net CPGE current is equal to their sum, \( I_{CPGE} = (I_{up} + I_{down}) \). Emptying of electrons in \( e_2 \) through interactions with optical phonons quickly quenches any current there. If \( h_0 \omega < E_{21} \) (LHS) the dominant transition (indicated by the thicker line) occurs between spin-up sub-bands at negative \( k_y \). Increasing \( h_0 \omega \) shifts this transition towards \( k_y = 0 \) (where \( h_0 \omega = E_{21} \)), eventually passing through \( k_y = 0 \) and onwards along positive \( k_y \), for \( h_0 \omega > E_{21} \) (RHS). The weaker transition moves in the opposite direction. This leads to a sign change in the current as the transitions pass through the minima of their respective bands (i.e. when \( \frac{\partial \alpha}{\partial k_y} \) changes sign), at an energy close to \( E_{21} \).

If the photon energy is continuously varied from \( h_0 \omega < E_{21} \) to \( h_0 \omega > E_{21} \), where \( E_{21} \) is also the separation of the \( e_1 \) and \( e_2 \) at \( k_y = 0 \) (see Figure 6.6), an important effect is observed in the generated CPGE current. As the energy is increased, the wavevector of carriers involved in both transition tends towards \( k_y = 0 \) so that conservation of energy may be observed. Just before \( k_y = 0 \) is reached, when \( h_0 \omega = E_{21} \), the two absorptions reach the bottom of their respective bands. There \( \frac{\partial \alpha}{\partial k_y} = 0 \), and \( I_{CPGE} \) reduces to zero. As \( h_0 \omega \) is increased still further, each transition continues to move into the opposite side of the spin sub-band from which it began. The values of \( \frac{\partial \alpha}{\partial k_y} \) in each case are thus reversed, and \( I_{CPGE} \) undergoes a sign change. The CPGE current is, therefore, approximately proportional to the differential of the sample’s absorption spectrum, crossing zero close to \( h_0 \omega = E_{21} \) where absorption in the well is a maximum. This sign-change is a unique characteristic of the CPGE, and is of particular interest to this work as a means of differentiating between currents due to the CPGE and SGE.
It may be noted that changing the circular-polarisation of the light from $\sigma^+$ to $\sigma^-$ in the above example also results in a sign change, as it reverses the direction of $\vec{\mathcal{E}}$. A sinusoidal variation of the CPGE current as a function of the helicity of the incident light in similar QW structures has been documented previously.[2,12,17].

### 6.2.3 The Spin Galvanic Effect

The spin-polarisation necessary for the SGE is illustrated in Figure 6.7. In the case of the experiment presented in this work, this will have been arrived at by photexcitation, like that described in section 6.2.2 for the CPGE effect. In fact, the experimental geometry will be assumed identical to that of Figure 6.5 in the following discussion. Therefore, many of the arguments in the context of the CPGE previously apply here also, including the requirement that any current (including that generated by the SGE) must flow along $k||k_y$. Immediately after excitation, the CPGE current is generated as detailed in section 6.2.2. However, momentum scattering interactions serve to thermalise the population imbalance in the $e_1$ sub-bands. On the timescale of such collisions, where $\tau_p$ is the momentum scattering time, the CPGE current will, therefore, decay exponentially. Once this thermalisation process is completed, the carrier occupation in each sub-band will be ‘flat’ (i.e. all carriers occupy the lowest energy states permitted by the Pauli Exclusion Principle), and the CPGE current will be negligible. However, there will still be a significant difference in the occupancy between each $e_1$ sub-band. This is because the spin relaxation time, $\tau_s$ (i.e. the timescale on which individual spins flip between spin sub-bands), and electron-hole recombination time, $\tau_R$, are both significantly longer than $\tau_p$. The resulting band occupation is illustrated in Figure 6.7 (RHS) at some time, $t$ (where $\tau_s>t>\tau_p$), after the initial excitation.
First Observation of the Spin-Galvanic Effect in Zero-Applied Magnetic Field

Figure 6.7 Schematic illustrating the microscopic origin of the SGE effect in an (001) grown QW for σ⁺ incident light with in-plane polarisation ̂σ⁺. Optical excitation generates a spin-polarisation ̂σ⁺. The population imbalance in spin sub-bands e₁ (the ground state) and e₂ (first excited state) giving rise to the CPGE is neutralised by momentum relaxation processes on timescale τₘ, leaving a flat-band occupation. At some time, t, later, asymmetric spin-flip processes 1 and 2 (occurring on timescale τₗ) generate a transient current (RHS); the SGE current. Note τₘ<τₗ. It may also be noted that τₘ<τ₉, so on a timescale comparable to τₙ spin-flip processes are primarily responsible for eroding the spin-polarisation in e₁, with the spin-down sub-band (larger population) emptying into the spin-up sub-band (smaller population) until an equilibrium is reached. There are two possible spin-flip transitions available to an electron in the spin-down e₁ sub-band; a transition to the RHS of the spin-up sub-band, and a transition to the LHS of the spin-up sub-band. The probability of a given transition is determined by the electron’s initial and final wavevectors, ̂kᵢ and ̂kᵣ respectively (where ̂kᵣ-̂kᵢ=Δk). For electrons of a particular energy (i.e. two electrons on either side of the spin-down sub-band) the available spin-flip transitions are illustrated in Figure 6.7 by solid and dashed arrows. Of the four transitions shown, two are clearly equal and opposite and, therefore, cancel one another out (solid lines). However, the remaining two (labelled 1 and 2 in Figure 6.7) are asymmetric (i.e. Δk₁<Δk₂), so one can reasonably expect their rates to be different (dotted lines).[1] This asymmetry in spin relaxation allows a transient imbalance to arise in each of the
e sub-bands, leading to currents $I_{\text{up}}$ and $I_{\text{down}}$ in the spin-up and spin-down sub-bands respectively. This process is illustrated in Figure 6.8. Both currents have the same direction and magnitude, so that a net transient current $I_{\text{SGE}} = I_{\text{up}} + I_{\text{down}}$ (where both have the same sign) flows along the [110] direction (i.e. perpendicular to $\hat{e}_x$). The magnitude of the SGE current is proportional to the difference in the rates of spin-flip transitions 1 and 2.

It should be noted that the SGE current does not exhibit a hard switch-on when the flat-band occupation of Figure 6.7 is reached. The asymmetric spin-flip transitions may occur immediately after the initial excitation, simultaneous with the thermalisation processes and, hence, the CPGE. As in the case of the CPGE, efficient emptying of the $e_2$ sub-bands via transitions involving optical phonons rapidly neutralises the spin-polarisation there. The contribution to $I_{\text{SGE}}$ from $e_2$ can, therefore, be neglected also.

Unlike the CPGE current, the magnitude of the SGE current is independent of the wavevectors of the initial exciting transitions (i.e. $\pm k_y$ in Figure 6.7). All memory of that wavevector is lost during the thermalisation phase preceding Figure 6.8. Rather the SGE current is dependent only on the intensity of the initial excitation. For this reason no sign change is observed in $I_{\text{SGE}}$ when $E$ is scanned through $E_{21}$, unlike $I_{\text{CPGE}}$.

Finally, whereas the CPGE relies on asymmetric (spin-dependent) optical excitation (i.e. preferentially emptying electronic states with particular wavevectors), a spin polarisation conducive for the SGE (i.e. the flat bands of Figure 6.7) may be generated by a variety of means, including electrical injection.
Figure 6.8 Schematic emphasising the transient nature of $I_{SGE}$ and its origins in asymmetric spin-flipping. In a) there is flat-band occupation, so no SGE current, $I_{SGE}$, flows. In b) after some time $\Delta t - \tau_{s1}$, spin-flip transition 1 occurs. This transition imbalances both spin-up and spin-down $e_i$ sub-bands, generating net $I_{SGE} = I_{up} - I_{down}$. Both component currents flow in the same direction and have the same magnitude. In c) after some further interval $\Delta t - \tau_{s2}$, spin-flip transition 2 occurs, re-balancing the bands and returning the system to flat-band occupation. Note the current does not periodically switch on and off as implied in the simplified example presented here. The spin-flip processes are statistical in nature, and the size of the current is determined by the difference in their rates.
6.3 Experiment

6.3.1 Samples

The GaAs/AlGaAs QW sample used in this experiment was grown in the (100) plane by MBE. The sample contained 30 $n$-type QWs, each 8.2nm wide with a room temperature carrier concentration of $n \sim 2 \times 10^{12} \text{cm}^{-2}$. The energy of the $e_1e_2$ intraband transition in the wells had been previously determined by FTIR absorption measurements to be $\sim 130 \text{meV}$, corresponding to a wavelength of $\sim 9.5 \mu\text{m}$. The sample was quadratic in shape, with Ohmic contacts placed in the centre of each of its four sides and at each of its corners. The current could thus be measured along four different crystallographic directions; [110], [1-10], [100], and [010].

6.3.2 Experimental Arrangement

These measurements were performed at the FELIX facility, the details of which can be found in section 1.2.2. The sample was placed in the beam at an oblique angle with respect to the FELIX beam’s angle of incidence in order to generate an in-plane component of spin polarisation. The FELIX macropulses were $\sim 5 \mu\text{s}$ long, comprised of $\sim 3 \text{ps}$ micropulses arriving $\sim 40 \text{ns}$ apart. The macropulse average power output at $\sim 10 \mu\text{m}$ is $\sim 20 \text{kw}$, delivered with a rep rate of $\sim 5 \text{Hz}$ (average power $\sim 0.1 \text{MW}$).

The (initially) linearly polarised FELIX beam was first passed through a Fresnel Rhomb in order to obtain circular polarisation. The sample was then illuminated, with the in-plane component of the polarisation vector aligned either $\hat{e}_x || [1-10]$ or $\hat{e}_y [110]$. The measured current was always perpendicular to the in-plane polarisation vector, as expected for both the CPGE and SGE. Current signals were amplified ($\times 5000$), before being passed to a boxcar integration module where a time gate triggered on the FELIX macropulses was applied. This helped minimise any spurious noise signals originating in the dark periods between FELIX pulses. The rise time of this module was $\sim 1 \mu\text{s}$, with its dc output representing the integrated current within the applied gate period (i.e. averaging was performed over $\sim 1$ macropulse). These values were read directly from an oscilloscope and recorded as the current values.
By scanning the FELIX wavelength through the $e_1e_2$ transition energy (i.e. from $\sim 8\mu m$ to $\sim 13\mu m$ in steps of $\sim 0.2\mu m$) measurements of the generated current as a function of photon energy were made. As will be shown in the following section, the CPGE and SGE currents could then be unambiguously differentiated by virtue of their different energy dependences (recall the CPGE has a zero crossing at $E_{21}$, but the SGE current does not).

The current signal's lineshape was assumed to be independent of wavelength (i.e. the gate width was not altered between readings), with FELIX pulse intensities remaining approximately constant as a function of wavelength. A schematic of the experimental set up is shown in Figure 6.9.
Figure 6.9 Schematic illustrating the experimental arrangement for generation and detection of spin-galvanic currents at FELIX. A circular polarised beam was achieved by passing the linearly polarised FELIX beam through a Fresnel Rhomb. The beam was then focussed onto the sample. Any current signal was then amplified, and gated using a boxcar integration unit triggered on the FELIX macropulses. The current was then read as the dc output of the boxcar module directly from the oscilloscope. The generated current was measured along $k_y\|[110]$ and $k_x\|[1-10]$ as a function of FELIX wavelength (between 8μm and 13μm).
6.3.3 Results and Discussion

![Graph showing photocurrents measured in [110] and [1-10] directions in GaAs QWs of 8.2nm width.](image)

**Figure 6.10** Photocurrents measured in the [110] and [1-10] directions in GaAs QWs of 8.2nm width. In each case incident radiation was perpendicular to the current. Currents are normalised by FELIX beam power -20kW per macropulse.

The measured photocurrents along [1-10] and [110] as a function of excitation energy are plotted in Figure 6.10. The in-plane component of the FELIX beam (polarised $\sigma^+$) was along $\hat{e}_x[1-10]$ and $\hat{e}_y[110]$ respectively. To explain the non-equivalence of the currents measured for each orientation, one can refer to a phenomenological description of their origins.[18] To that end, one can write the following expressions for $I_{SGE}$ and $I_{CPGE}$ in the x- and y-directions:

$$I_{SGE} = \ldots$$
$$I_{CPGE} = \ldots$$
First Observation of the Spin-Galvanic Effect In Zero-Applied Magnetic Field

\[ I_{SGE-x} = Q_{xy} S_y, \quad I_{SGE-y} = Q_{yx} S_x \]  \hspace{1cm} (6.10)

\[ I_{CPGE-x} = R_{xy} \hat{e}_x E_0^2 P_{circ}, \quad I_{CPGE-y} = R_{yx} \hat{e}_y E_0^2 P_{circ} \]  \hspace{1cm} (6.11)

where \( Q \) and \( R \) are second rank pseudotensors whose elements are dependent on the structural geometry of the QW (i.e. the details of any BIA, IIA and SIA). \( S \) is the average spin polarisation in the QW, \( E_0 \) is the intensity of the incident electromagnetic wave, and \( P_{circ} \) is the helicity of that radiation. The formulation of these expressions emphasises the fact that the CPGE current relies on the photo-excitation of carriers, while generation of the spin polarisation required for the SGE current does not. However, for these measurements the photo-excitation was used in both cases, so \( S \) is actually proportional to \( E_0 \hat{e} P_{circ} \) also.

The spin-orbit coupling term of the Hamiltonian, \( \mathcal{H}_{so} \), may then be written as:

\[ \mathcal{H}_{so} = \sum_{lm} \gamma_{lm} \sigma_l \sigma_m \]  \hspace{1cm} (6.12)

where \( \gamma \) is a second rank pseudotensor proportional to \( Q \) and \( R \), with constants of proportionality are \( A_{SGE} \) and \( A_{CPGE} \) respectively. It may be noted that equation 6.12 is equivalent to equation 6.4, only using a more general tensor notation. For \( C_{2v} \) symmetry with \( k_s || 1-10 \) and \( k_y || 110 \) (note the rotation in coordinate axis relative that used in section 6.2.1), it can be shown only two elements of \( \gamma \) are non-zero: \( \gamma_{xy} \) and \( \gamma_{yx} \), where \( \gamma_{xy} \neq \gamma_{yx} \). The fact that only the off-diagonal elements (or cross-terms) are non-zero expresses the fact that the spin-polarisation and current are always perpendicular to one another. Furthermore, as \( \gamma_{yx} \) describes the interaction with spins aligned along \( k_s || 1-10 \) (see equation 6.10), by comparing equation 6.4 and equation 6.12 it can be shown \( \gamma_{yx} \) is proportional to \( (\alpha + \beta) \). Similarly, \( \gamma_{xy} \) gives the interaction with spins aligned along \( k_y || 110 \), and is proportional to \( (\alpha - \beta) \). It is therefore possible to express \( \alpha \) and \( \beta \) in terms of \( \gamma_{xy} \) and \( \gamma_{yx} \) as follows:

\[ \alpha = \gamma_{SIA}^{(v)} = \frac{\gamma_{xy}^{(v)} - \gamma_{yx}^{(v)}}{2} \]  \hspace{1cm} (6.13)
\[ \beta = \gamma_{\text{BIA}}^{(v)} = \frac{\left(\gamma_{xy}^{(v)} + \gamma_{yx}^{(v)}\right)}{2} \]  \hspace{1cm} (6.14)

where \( \nu = 1, 2 \) refers to the \( e_1 \) and \( e_2 \) sub-bands respectively. Combining equations 6.13 and 6.14 with equation 6.11, the net currents along the [1-10] and [110] may, therefore, be written as the sum of contributions from the CPGE and SGE as follows:

\[ I_x = A_{\text{SGE}} \left(\gamma_{\text{SIA}}^{(1)} - \gamma_{\text{BIA}}^{(1)}\right) S_y + A_{\text{CPGE}} \left[\left(\gamma_{\text{SIA}}^{(1)} - \gamma_{\text{BIA}}^{(1)}\right) - \left(\gamma_{\text{SIA}}^{(2)} - \gamma_{\text{BIA}}^{(2)}\right)\right] E_0 \hat{e} \hat{P}_{\text{circ}} \]  \hspace{1cm} (6.15)

\[ I_y = A_{\text{SGE}} \left(\gamma_{\text{SIA}}^{(1)} + \gamma_{\text{BIA}}^{(1)}\right) S_x + A_{\text{CPGE}} \left[\left(\gamma_{\text{SIA}}^{(1)} + \gamma_{\text{BIA}}^{(1)}\right) - \left(\gamma_{\text{SIA}}^{(2)} + \gamma_{\text{BIA}}^{(2)}\right)\right] E_0 \hat{e} \hat{P}_{\text{circ}} \]  \hspace{1cm} (6.16)

where terms related to the \( e_2 \) sub-bands appear in the CPGE current component due to the effects dependence on the wavevectors at which the photo-excitation occurs. More specifically, the CPGE current is dependent on those wavevectors, which are themselves dependent on the energy separation of the relevant states in the \( e_1 \) and \( e_2 \) sub-bands, which is in turn dependent on the magnitude of the spin-splitting in both the \( e_1 \) and \( e_2 \) sub-bands.

Using equations 6.15 and 6.16, it is thus possible to perform a qualitative analysis of the results shown in Figure 6.10. Along \( k_x || [1-10] \) the current exhibits a clear sign change close to \( E_{21} \), indicating the dominance of the CPGE current (see section 6.2.2). This result is reasonable based on inspection of equation 6.15, as the amplitude of the SGE current is proportional to the difference between \( \alpha \) and \( \beta \). Note that the CPGE contribution contains similar cancelling terms, but the additional dependence on the (larger) \( e_2 \) sub-band splitting serves to pacify the effect somewhat. In the [110] direction no sign change is observed, implying the dominance of the SGE current (see section 6.2.3). This is consistent with the prediction of equation 6.16, where it is the sum of \( \alpha \) and \( \beta \) which is relevant. Therefore, the suppression of the SGE current is lifted along [110], while the simultaneous CPGE current is relatively weak due to its dependence on the difference between the splitting in the \( e_1 \) and \( e_2 \) sub-bands. It may be noted that in QWs possessing \( C_{2v} \) symmetry, the CPGE current is always expected to reduce to zero at energies close to \( E_{21} \). Therefore, the current observed along [110] at such energies (i.e. \( \sim 130 \text{meV} \)) is a pure SGE current with no CPGE component.
It may be noted that this measurement has not fully differentiated the CPGE and SGE as both are, in general, present in both [1-10] and [110] directions. However, it has been possible to make an unambiguous determination of the dominant current mechanism in each direction without applying a magnetic field. In particular, the current along [1-10] has been identified as being due mainly to the SGE.

### 6.3.4 Conclusions and Future Work

The anisotropic nature of the linear spin-splitting in the presence of both BIA and SIA has been exploited as a means of suppressing the CPGE effect in n-type GaAs QW’s. Making use of the tuneability of FELIX, the behaviour of the currents along the non-equivalent [110] and [1-10] directions has been studied as function of photon energy. Along [110] significant suppression of the CPGE current has been observed in agreement with the predictions of a phenomenological model. The dominant current in this direction was, therefore, determined to be due to the SGE. For photon energies close to the energy separation between the ground state and first-excited state in the QW the CPGE current is expected to be negligible along all directions, so the non-zero current measured along [110] at such energies is a pure SGE current. The SGE current is, therefore, unambiguously identified without application of a magnetic field for the first time.

In future, using a tuneable infrared laser such as FELIX it should be possible to choose the lasing wavelength to be resonant with the CPGE zero-crossing energy, thereby allowing the SGE current to be studied as a function of various material parameters in (near) isolation. A better understanding of the SGE current may help in developing a device which operates in reverse, possibly leading to class of devices capable of generating a spin-polarisation via current injection.

Finally, mention should be made of an interesting effect which occurs in zinc-blende QW’s when $\alpha$ and $\beta$ are tuned to be equal. As an example, consider the (001) grown structures used in this work. When $\alpha=\beta$, the effective magnetic fields (and hence the spin-eigenstates) are found to be independent of wavevector, uniformly aligning along [1-10]. As a result of this isotropic alignment, any spin with a component parallel to
[1-10] experiences a suppression of D'yakonov-Perel (DP) spin relaxation (see section 5.2.2.1), effectively extending the spin lifetime in such structures. Recently it has been suggested that for (110) and (111) grown zinc-blende QW's it may be possible to achieve DP suppression for more than just one component of spin, extending it to two or even all three components simultaneously.[19] This has important implications for the development of spintronic devices, including longer drift lengths and spin-storage (or memory). The resonance at $\alpha=\beta$ may also be employed as a means of controlling the polarity of a spin-polarised current. That is, an applied electric field perpendicular to the QW plane can change $\alpha$, de-tuning the $\alpha=\beta$ resonance, in turn altering the spin-eigenstates and instigating a precession of [1-10] aligned spins. In that case, the magnitude of the electric field also determines the magnitude of the effective magnetic field, thereby allowing the precession rate to be controlled, as required in proposed spin-FET designs (see section 5.1.1).[20]

However, creating successful devices utilising $\alpha$ and $\beta$ in this way hinges on the ability to design a structure where the parameters are appropriately tuned. It has been demonstrated by Ganichev et al that by measuring the SGE current along several directions the ratio $\alpha/\beta$ can be determined.[4] These measurements were performed utilising applied magnetic fields to differentiate SGE and CPGE currents, although the measurement performed at FELIX demonstrates an alternative method for studying the SGE without the application of such a field.

6.4 References


First Observation of the Spin-Galvanic Effect in Zero-Applied Magnetic Field


7 The Effect of Pressure on the Radiative Efficiency of InAs-Based Devices

7.1 Introduction

Interest in the development of devices operating in the 2-5\,\mu m spectral region has grown considerably in recent years, largely due to the fact that the rotational and vibrational modes of many common gases are found at such wavelengths. The main aim of research in this area is, therefore, to engineer a device able to generate sufficient output power for use in a selective, highly efficient optical gas sensor, capable of identifying target gases such as CH$_4$, CO$_2$ and CO. Presently, heterostructures based on alloys of InAs system are believed to be the best candidates for successfully engineering such an MIR emitting light source.[1] The main problems associated with fabricating devices based on NGSs (like InAs) is reducing unwanted non-radiative loss processes, such as Auger recombination and defect scattering.[2,3,4] Hydrostatic pressure has been extensively used to study recombination mechanisms in semiconductor laser devices. However, until recently such measurements have been limited to the wavelengths of interest to telecommunications applications (i.e. 1.3-1.5\,\mu m).[5] In order to adapt these techniques for MIR wavelengths several fundamental changes to the experimental arrangement were necessary; a subject of further discussion in section 7.3.2.

The heterostructure devices provided for this work by Lancaster University were grown on InAs substrates by liquid phase epitaxy (LPE). A gettering technique was
also employed during growth, involving adding ytterbium (Yb) to the InAs melt in an effort to reduce the residual carrier concentration of the resulting devices.[6] Post-growth, the devices were processed into mesa etched LEDs using standard photolithographic techniques. Emission at 3.3μm was observed from both a type-I laser structure and a type-II LED grown in this way, with the light output vs injected current (or L-I) characteristic of each measured as a function of hydrostatic pressure using a helium gas compressor.

7.2 Theory

7.2.1 Auger Recombination

There are several mechanisms by which an electron and a hole may recombine in a semiconductor (device) without emitting a photon (at the wavelength of interest). These non-radiative mechanisms include trapping at mono-molecular defects, surface recombination, Auger recombination, and current leakage (in heterostructures). In conventional mid-infrared inter-band devices, Auger recombination is generally believed to be the most efficient of these mechanisms. Minimising such processes is, therefore, a key engineering goal.

Auger recombination is a scattering process involving three charge carriers. Several variations exist, and these may be classified into one of two categories; direct band-to-band or phonon assisted.[7,8] A selection of the most common Auger recombination processes are illustrated in Figure 7.1. Each variation is labelled according to the bands occupied by the scattering carriers. For example, the CHCC process involves an indirect recombination between an electron in the Conduction band and a hole in the Heavy hole band (hence CH), with momentum and energy conservation maintained by the excitation of a second electron already in the conduction band to a higher energy state in the Conduction band (hence CC). The three carriers have in effect undergone an inelastic collision, redistributing their energies amongst themselves. A phonon-assisted transition is identical to a direct process, only in the former case the carrier-carrier interactions leave the system in an intermediate state that does not conserve momentum. The ‘missing’ momentum instead comes from a lattice scattering process (i.e. a phonon).
The Effect of Pressure on the Radiative Efficiency of InAs-Based Devices

Figure 7.1 Schematics illustrating some of the most common Auger recombination processes in bulk III-V semiconductors. All involve three charge carriers; either two electrons and one hole (e.g. CHCC), or one electron and two holes (e.g. CHLH). The names of each process denote the band the participating carriers occupy. Note CHCCP is a phonon-assisted process, and involves an additional interaction between a heavy hole and a lattice phonon. All Auger recombination processes may have phonon-assisted variants.

Because Auger recombination involves three carriers, two of one type and one of the other, the recombination rate, $R_{\text{Auger}}$, can be expressed as:

$$R_{\text{Auger}} = Cn^2p,$$

or

$$R_{\text{Auger}} = Cp^2n$$

(7.1)
The Effect of Pressure on the Radiative Efficiency of InAs-Based Devices

where \( C \) is the Auger recombination coefficient, and \( n \) and \( p \) are the electron and hole densities respectively. For the sake of simplicity low-doping levels will be assumed such that the approximation \( np \) holds. Equation 7.1 may then be redefined as follows:

\[
R_{\text{Auger}} = Cn^3 \tag{7.2}
\]

In this regime Boltzmann statistics should also hold as an approximation, and in analogy with radiative recombination (as discussed in section 2.1.1) one may write:[9]

\[
C = C_0 \exp\left(-\frac{E_{\text{Auger}}}{k_B T}\right) \tag{7.3}
\]

Where \( C_0 \) is another coefficient analogous to the Einstein coefficient (only in this case as applied to scattering processes), and \( E_{\text{Auger}} \) is the Auger activation energy. The latter arises from the geometric limitations imposed by the size of the bandgap and curvature of the bands (as they appear in an \( E-k \) diagram) if both momentum and energy are to be conserved in an Auger process. For the CHCC process, \( E_{\text{Auger}} \) is given by:[10]

\[
E_{\text{Auger}}(\text{CHCC}) = \frac{m_e}{m_e + m_h} E_g \tag{7.4}
\]

where \( m_e \) and \( m_h \) are the conduction band and heavy hole effective masses respectively. It can be seen from equation 7.4 that decreasing the bandgap also decreases the Auger activation energy. Further, for a smaller bandgap the intrinsic carrier densities at a given temperature can be expected to be higher (i.e. \( E_g/k_B T \) is smaller). Together, these points explain why Auger recombination presents such a problem for NGSs, particularly when they are incorporated in devices whose emission is reliant on inter-band radiative transitions. It may also be noted that \( E_{\text{Auger}}(\text{CHCC}) \) is minimised if the effective masses of conduction band and heavy hole band are equal. A similar expression may be written for other Auger processes also by substituting the
relevant band masses and energy gaps. For example, for the CHSH process $E_{\text{Aug}}$ is given by:

$$E_{\text{Aug}}(\text{CHSH}) = \frac{m_{so}(E_g - \Delta_{so})}{(2m_k + m_e - m_{so})} \quad (7.5)$$

### 7.2.2 High Pressure

Although temperature remains the most readily manipulated and, therefore, most commonly used thermodynamic variable, the application of high pressure is an invaluable diagnostic tool for semiconductor device physics.\[11\] Several reviews have been published on the subject, where the techniques' still largely untapped potential is extolled.\[12,13,14\] Application of hydrostatic pressure in particular has led to the identification of the dominant loss mechanism in quantum well lasers.\[15\]

Although the changes that occur in a semiconductors bandstructure under applied hydrostatic pressure are extremely complex, there is a simple qualitative argument which can be used to explain the general trends observed. When a semiconductor is compressed, its constituent atoms are forced closer together. This decreases their bond lengths, resulting in a greater overlap of their atomic orbitals, increasing their interaction energy. In a III-V semiconductor, the valence band and conduction band are comprised of $p$-like bonding and $s$-like anti-bonding orbitals respectively, so the interaction between the two is repulsive. An increase in the interaction energy pushes the valence and conduction bands farther apart in energy, therefore, and hydrostatic pressure can be expected to increase such a materials bandgap. As a result, processes which are dependent on the bandgap energy, like Auger recombination, can be studied. Specifically, the power output from an emitter should change as a function of pressure, due to the shifting balance between radiative and non-radiative recombination mechanisms\[19\], thereby allowing their relative rates to be extracted.

When hydrostatic pressure is applied to a III-V semiconductor or alloy, the direct bandgap at the Γ-point increases by around $\sim 10 \text{meV/kbar}^{-1}$, while at the X-point a

\[19\] The dependences of Auger recombination and radiative recombination on the bandgap are not the same. This can be seen by comparing equation 7.4 and equation 2.7.
decrease of approximately $\sim 1 \text{meV/kbar}^1$ is observed, and at the L-point there is an increase of $\sim 5 \text{meV/kbar}^1$.[19]

### 7.2.3 Heterostructure Type

The heterostructures studied in this work are of type-I and type-II design. The former has simple well-like band alignment for both electrons and holes, leading to the confinement of both within the same region of the device. The band offsets of a type-II structure, however, are designed to confine electrons and holes in different regions of the device. Recombination can then occur only when an electron and/or hole tunnels through the potential barrier that separates them. By keeping the carriers spatially separated in this way, it has been suggested Auger recombination can be suppressed.[16,17,18] Examples of type-I and type-II band alignments are illustrated in Figure 7.2.

**Figure 7.2** Schematic of the band profiles for basic type-I and type-II alignments. The bandgap of material A is smaller than material B, with positive band offsets for both conduction and valence bands. This is a type-I alignment. Material C is also a smaller bandgap than material D, but the conduction band of material C is at higher energy (i.e. a negative offset). Electrons are therefore confined to the 'barrier' material D, while holes are confined to material C. This a type-II offset. For materials E and F this alignment is reversed, though it is also a type-II alignment. In type-II heterostructures, recombination can only occur between carriers where their wavefunctions extend into the (respective) barriers due to quantum tunnelling.
7.3 Experiment

7.3.1 Samples

The type-I laser device studied was designed to operate at 3.3µm at room temperature, although it was subsequently found to only operate at low temperatures. Pressure measurements were, therefore, limited to below the lasing threshold (i.e., spontaneous emission). The device was a double-heterostructure (DH) design, comprised of InAs/InAs$_{0.94}$Sb$_{0.06}$/InAs layers. The very small conduction band offset of only ~20meV is actually smaller than $k_B T$ (~26meV), implying electrons should be able to make it into the barrier layers and contribute a weak type-II component to the recombination.

The type-II LED device studied is also a DH design. This consisted of an unintentionally doped, 0.7µm thick n-InAs active layer sandwiched between p- and n-type InAsSbP confinement layers. The device structure is illustrated in Figure 7.3. The phosphorous content in the confinement layers was 0.40, giving $E_g = $570meV under ambient conditions. This results in a conduction band offset of around ~215meV, also under ambient conditions, considerably larger than in the type-I device, thus improving carrier confinement. There is no structural potential well in these devices, with confinement believed to be provided by band-bending at the heterostructure interface (a consequence of electrostatic repulsion). Due to the nature of the band-bending, electrons are expected to be confined on the InAs side of the interface and the holes on the InAsSbP side, thereby leading to the type-II emission. This band profile is illustrated in Figure 7.4, where the radiative transition occurring between confined states is also indicated. Due to the electrons smaller effective mass, their tunnelling into the InAsSbP layer should dominate over holes moving in the opposite direction, so the majority of radiative recombination should occur in the InAsSbP layer.
An Electroluminescence (EL) measurement performed at Lancaster shows a single emission peak at 3.28\(\mu\)m, as shown in Figure 7.4 (RHS, larger than the bandgap of the n-InAs active region. It is believed that emission corresponds well with type-II recombination from confined states at the InAs/InAsSbP interface, thus supporting the argument for type-II alignment.

\[ \begin{align*}
\text{InAsSbP} & \quad E_g = 0.570\text{eV} \\
& \quad \lambda_g = 2.20\mu\text{m} \\
\text{Interface} & \quad E_g = 0.378\text{eV} \\
& \quad \lambda_g = 3.28\mu\text{m} \\
\text{InAs} & \quad E_g = 0.354\text{eV} \\
& \quad \lambda_g = 3.50\mu\text{m}
\end{align*} \]

**Figure 7.4** The emission characteristics of the type-II LED structure. On the LHS a schematic of the band-bending at the interface between InAsSbP and InAs layers is shown. On the RHS an electroluminescence spectrum from the device is shown. This was measured at Lancaster University at room temperature, with a constant drive current of 50mA at 430Hz and 50% duty cycle. The transition leading to the 3.28\(\mu\)m emission is indicated on the LHS schematic, where it is shown to be consistent with transitions between quantised states at the type-II interface.
In order to couple the light from a semiconductor into a detector whilst under applied pressure, it is imperative that the pressure medium be transparent to the emitted radiation. At NIR wavelengths, liquids are typically used (usually oil or alcohol). However, at MIR wavelengths these media are not suitable, as the emitted light would be absorbed by vibrational bands in the liquid. A solution to this problem is to use a gaseous pressure medium, of which there several candidates transparent at MIR wavelengths (e.g. He, N, etc.). Beyond transparency there are two further advantages of using a gas as the pressure medium. Firstly, the pressure medium (and hence the sample) can be cooled far below room temperature without the effects of non-hydrostatic behaviour (e.g. freezing). Pressure studies as a function of temperature are, therefore, possible with a gas-based system (though this ability is not used in this work). Secondly, when using a liquid pressure medium the act of applying pressure changes its refractive index, which in turn alters the reflectivity of the device facet and, hence, its light output. However, the refractive index of a gas remains approximately constant over a broad range of pressures, alleviating this problem somewhat.

For this study a compressor based on a helium pressure medium was used. This system was designed and manufactured at the High Pressure Research Centre at the Polish Academy of Sciences. Capable of reaching pressures as high as 1500MPa, the system consists of three stages. Each consecutive stage has an increasing compression ratio which serves to amplify the pressure in any of the previous stages (or for the first stage is equal to the driving oil pressure). Further information about the pressures achievable at each stage is provided in Table 7.1.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Pressure Range</th>
<th>Pressure (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial gas pressure</td>
<td>8-15MPa</td>
<td></td>
</tr>
<tr>
<td>Maximum hydraulic pressure</td>
<td>70MPa</td>
<td></td>
</tr>
<tr>
<td>Stage I pressure range</td>
<td>0-70MPa</td>
<td></td>
</tr>
<tr>
<td>Stage II pressure range</td>
<td>0-350MPa</td>
<td></td>
</tr>
<tr>
<td>Stage III pressure range</td>
<td>0-1500MPa</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.1 Pressures attainable at each stage in the helium compressor
The Effect of Pressure on the Radiative Efficiency of InAs-Based Devices

7.3.3 The Pressure Cell

The GOC 102 pressure cell (also purchased from the High Pressure Research Centre, Polish Academy of Sciences) is designed for optical studies of solids under high gas pressures up to 1200MPa. It is constructed from BeCu, a non-magnetic material which permits use in high magnetic fields. A long capillary extends from the cell to the compressor as a means of transferring the He gas from one to the other, and is also made from BeCu. The outer diameter of the capillary is 3mm while the internal diameter is restricted to just 0.3mm. At the compressor end of the capillary is a manganin pressure gauge, the resistance of which varies with pressure, thereby providing a means of monitoring the pressure. However, a more accurate measurement of the pressure applied to the sample is available at the cell end by means of a second gauge, consisting of a pressure sensitive semiconductor (SPG10) situated within the sample space itself. This gauge is a single crystal of heavily doped bulk $n$-type InSb, and exhibits a high sensitivity to hydrostatic pressure, but a low sensitivity to non-hydrostatic stress.

![Figure 7.5](image_url) The GOC 102 Gas Optical Cell (not to scale). Pressurised He enters the cell through the capillary at the top. On the left is the electrical plug with access to the wires attached to the sample mount. On the right is the optical plug with sapphire window. The sample to be studied sits on the green platform atop the sample mount inside the sample space, approximately ~10mm from the sapphire window.
The pressure cell is depicted schematically in Figure 7.5. At either end of the cell body is a socket into which either an optical or electrical plug can be inserted. The optical plug consists of a sapphire window 6mm in diameter and 6mm thick, mounted on top of a 2mm diameter exit port. Sapphire is used because it is both able to withstand high pressures without cracking, and is transparent at MIR wavelengths (~85% transmittance between 0.3μm and 5μm).

The electronic plug feeds twelve wires (eleven Cu and one constantin) into the cell, with a pressure seal maintained through use of a shaped conical plug, which makes a seal with a similarly conical recess in the cell’s body. This recess is usually filled with powder in which the wires are buried. Of the twelve wires, two (one constantin and one Cu) are used to form a thermocouple, intended for measuring the temperature in the sample space. Four more are used to connect to the SPG10 pressure gauge. The resistance of the gauge is well below 0.1Ω requiring, therefore, the application of a four terminal method to measure the resistance. The remaining six wires can be used for any purpose, and in this work some were used to power the devices placed inside the pressure cell.

7.3.4 Experimental Procedure

The L-I characteristic of each sample has been measured as function of hydrostatic pressure. After passing through the sapphire end window of the pressure cell, the emitted radiation was measured using a nitrogen-cooled InSb detector. The signal was then processed using phase sensitive techniques. This arrangement is illustrated schematically in Figure 7.6.

Note that the emission was observed through the n-type (top) side of the samples, which was 300μm in diameter and had a ring-like contact. This detail allows inter-valence band absorption (IVBA) due to the light passing through the device itself to be neglected, as that process is significant only in p-type materials.
The Effect of Pressure on the Radiative Efficiency of InAs-Based Devices

Figure 7.6 The experimental arrangement used to measure the light output of the MIR InAs-based LEDs as a function of pressure. It was necessary for the compressor and pressure cell to be isolated from the rest of the laboratory by means of a protective enclosure for safety reasons.

7.3.5 Results and Discussion

The room temperature integrated spontaneous emission, \( L_{spon} \), from the type-I laser at a constant driving current \( (I=2A) \), for pressures between 0kbar and 11.2kbar, is shown in Figure 7.7. An increase in output of only a factor of \(~1.87\) was observed over the pressure range.

Figure 7.7 Room temperature spontaneous emission from the type-I laser sample at pressures between 0kbar and 11.1kbar. The drive current was a constant 2A; below the lasing threshold. An increase in light output of \(~1.87\) was observed over the pressure range.
In Figure 7.8, $L_{\text{spon}}$ from the type-II device as a function of applied current, for between 0kbar to 9.4kbar, is shown. The variation in detector response due to the blue-shift of the emission with applied pressure has been corrected for. Focussing on just one applied current, 140mA (see the inset plot in Figure 7.8), a factor of \(\sim 3.74\) increase in the light output has been observed.

Figure 7.8 Room temperature spontaneous emission from the type-II LED sample for pressures between 0kbar and 9.4kbar as function of drive current. The slope of the L-I curves is seen to increase incrementally (and consistently) with applied pressure. Shown inset is the spontaneous emission as function of pressure for a constant drive current of 140mA. An increase in light output of \(\sim 3.74\) was observed over the pressure range.
The current density, $I$, driven through the device can be expressed as follows:

$$ I = I_{\text{leak}} + ed_{\text{eff}} \left( An + Bn^2 + Cn^3 \right) $$  \hspace{1cm} (7.6)

where $I_{\text{leak}}$ is the leakage current, $e$ is the electronic charge, $n$ ($=p$) is the carrier density, and $A$, $B$, and $C$ are the coefficients for mono-molecular (or defect) scattering, radiative recombination, and Auger recombination respectively. It is assumed that $C$ is the sum of all possible Auger recombination processes. Due to the large contact area of the device, any leakage current should be limited to carriers making it over the barrier at the heterojunction interface. However, the barriers are sufficiently high that that process should be very small, so the leakage current is not expected to be a significant current path in these devices. This conclusion is supported by the EL spectrum shown in Figure 7.4, which clearly shows only one peak, implying emission at only one wavelength (corresponding to recombination at the interface). Furthermore, it has been estimated for this structure that at a pressure of 9.4kbars the conduction band and valence band offset corrections are only 4meV and 1meV respectively. This is only a tiny fraction of the full offset, so it is also assumed that the leakage current is approximately independent of pressure (within the measured range), and $I_{\text{leak}}$ can be neglected in equation 7.6.

The Yb gettering technique applied during growth has produced a very low value for the residual carrier concentration in the active region, $n<5\times10^{15}\text{cm}^{-3}$. The $An$ term in equation 7.6 is therefore estimated to be very small at ambient conditions (this conclusion has been confirmed elsewhere).[1] The capture cross section for monomolecular recombination is also approximately independent of pressure. Hence, the $An$ term in equation 7.6 may also be neglected. For a constant drive current, $I$, equation 7.6 thus reduces to:

$$ ed_{\text{eff}} \left( Bn^2 + Cn^3 \right) = I $$  \hspace{1cm} (7.7)

This implies, therefore, a 1:1 trade off between radiative emission and Auger recombination, where any change in $Bn^2$ as a function of applied pressure is compensated for by an equal and opposite change in $Cn^3$. An upper limit for the the
radiative recombination coefficient can be found by assuming that Auger recombination is completely suppressed at the highest pressure, leading to:

\[ B_{\text{STP}}n^2 + C_{\text{STP}}n^3 = I = B_{\text{high}}n^2 + C_{\text{high}}n^3 \] (7.8)

where \( B_{\text{STP}} \) and \( C_{\text{STP}} \) are the ambient (i.e. atmospheric pressure) values of the respective coefficients, and \( B_{\text{high}} \) and \( C_{\text{high}} \) are the high pressure values. It is being assumed that \( C_{\text{high}}=0 \) and \( B_{\text{high}}n^2 = \alpha B_{\text{STP}}n^2 \) for the purposes of this analysis, which leads to:

\[ \frac{B_{\text{STP}}n^2}{B_{\text{STP}}n^2 + C_{\text{STP}}n^3} \leq \frac{1}{\alpha} \] (7.9)

The light output from the type-II LED increased by a factor \( \alpha=3.74 \) during the pressure increase from 0kbar to 9.4kbar, so according to equation 7.9 at most only 26.7% of the current passing through device is going into radiative emission at ambient pressure. It is thus confirmed that Auger recombination is the dominant current path in this device. In contrast, a similar analysis of the results for the type-I device (\( \alpha=1.87 \)) yields a radiative emission contribution at ambient pressure of 54%. Therefore, it is conceivable that radiative recombination is the dominant current path in that device.

This result is contrary to expectation, as it was hoped the type-II interface in the LED would reduce Auger recombination. However, an explanation can be offered if one considers the width of the effective active region, \( d_{\text{eff}} \). In the type-II device, carriers are localised near the interface, but because of the potential barrier spatially separated. Hence, the effective active region width is comparable to the distance through which carriers can tunnel, which for electrons is estimated to be \( d_{\text{eff}}=15\text{nm} \). It is believed, therefore, that although the Auger recombination coefficient \( C \) is reduced by the type-II interface design, the small effective active region which also results leads to a carrier concentration far larger than that in the type-I structure. The subsequent increase in \( n^3 \) is believed to be larger than the suppression achieved in \( C \) through use of the type-II structure resulting, therefore, in a larger net value for \( Cn^3 \) compared to
The Effect of Pressure on the Radiative Efficiency of InAs-Based Devices

the type-I structure. The stronger increase in light output as a function of pressure for the type-II device may then be interpreted as a suppression of the CHCC process due to the induced blueshift in $E_g$ (see equation 7.3).

For both type-I and type-II structures, the dominant Auger recombination channels are CHCC and CHSH (see Figure 7.1). CHLH and other valence band-related processes not involving the split-off band are important only in materials where the split-off energy, $\Delta_{so}$, is very large and the valence band mass is very light. In this case, the relative proximity of the split-off and light hole bands results in strong mixing, making the light hole mass very heavy. Regarding CHSH, as the pressure is increased it is expected that the bandgap should at some point come into resonance with split-off energy (i.e. $E_g = \Delta_{so}$), causing the CHSH process to go through a maximum. Therefore, radiative emission (i.e. $L_{spon}$) should therefore be greatly reduced at that energy. On the InAsSbP side of the junction, where recombination is believed to occur predominantly, the bandgap is smaller than the split-off energy at ambient conditions (i.e. $E_g < \Delta_{so}$), with $E_{g eff}=0.328$eV (from the EL emission peak) and $\Delta_{so} \approx 0.5$eV (it may be noted that in the literature there is a considerable uncertainty in $\Delta_{so}$). However, experimentally no resonant behaviour was observed in the light output of either device as function of pressure, indicating either the condition $E_g < \Delta_{so}$ holds over the full pressure range, or the coefficient for the CHSH process is simply very small. In any case, it appears CHSH is insignificant in both devices, leaving CHCC as the prime candidate for the dominant Auger recombination process.

More detailed theoretical calculations for the type-II device have been made by Dr. A. Andreev of the University of Surrey, using a type-II flat band quantum well model.[16] At ambient pressure, he has shown that $B=10^{-7}$cm$^3$s$^{-1}$ and $C=8 \times 10^{-27}$cm$^6$s$^{-1}$, corresponding to a radiative recombination contribution to the current density of 14% at $P=0$kbar and 40% at $P=9.4$kbar (for a constant current of 140mA). Hence, that somewhat simplified model predicts an increase in light output of a factor of 2.9 over that pressure range, in reasonable agreement with the experimentally determined value of 3.74.
7.3.6 Conclusions

The spontaneous emission of type-I and type-II InAs-based devices emitting at 3.3\textmu m have been studied as a function of pressure. An increase in the light output by both devices has been observed as a function of pressure, allowing a general comparison of their performances to be made in terms of the radiative and non-radiative processes. Auger recombination processes have been shown to provide the dominant current path in the type-II device, while it appears radiative and non-radiative recombination rates in the type-I device may be more comparable. The CHSH Auger process has been shown to be insignificant in both devices.

The comparatively high Auger recombination rate in the type-II structure is believed to be a result of the short tunnelling distances of carriers confined at the heterostructure interface, which leads to a small effective active region and, hence, a large carrier concentration. Therefore, although $C$ may be smaller in the type-II device as anticipated, $Cn^3$ is in fact larger than in the type-I device.

7.4 References


8 Conclusions

Several areas of material and device physics have been studied in this thesis. These have ranged from optical characterisation of mid-infrared materials, measurements of spin relaxation and dynamics in certain narrow gap semiconductors and a wide gap quantum well respectively, and investigations of recombination mechanisms in mid-infrared emitting InAs-based devices using high pressure. The following provides a summary of the conclusions arrived at as a result. Additional discussion, including any future work, can be found at the end of each chapter.

Using a step-scanning Michelson interferometer, the design and construction of which comprises part of this thesis’s work, photomodulation signals have been successfully measured in a range of material/heterostructures. At shorter wavelength (i.e. ~1.5-3.5μm), where the operational range of the Michelson interferometer-based arrangement and a pre-existing grating-based version overlap, good agreement has been obtained between the two available systems. At longer wavelength (i.e. ~5-6.5μm), beyond the reach of the grating-based arrangement, photomodulated transmittance has been measured from a pair of similar InSb epilayers. This has yielded bandgap energies consistent with expectation, although an anomalous evolution of the resulting features’ lineshapes has been observed as a function of temperature. The increased efficiency of Auger recombination processes (and their sensitivity to changes in bandgap) in narrow gap semiconductors is believed to be responsible for the latter effect. To the author’s knowledge, these measurements provide a new long wavelength benchmark for temperature dependent photomodulation measurements, the previous yardstick being photomodulated
reflectivity measurements on InAs at room temperature (i.e. \( \sim 3.5\mu m \)), performed using a grating spectrometer.

Photoluminescence has been used as means of unambiguously observing bandgap reduction in dilute nitride alloys based on narrow gap semiconductors: in particular GaSb\(_{1-x}N_x\) and InAs\(_{1-x}N_x\). Unfortunately, in the case of GaSbN no reduction in bandgap has been observed, with the majority of nitrogen containing samples yielding no photoluminescence or photomodulated reflectance signals whatsoever. Results consistent with the literature have been observed from binary control samples (i.e. GaSb), however, while those GaSbN samples which did provide signals appear (based on these measurements) to simply be lower quality GaSb. In contrast, similar studies of InAsN have proven quite successful, with an unambiguous reduction in bandgap observed in both low and high temperature photoluminescence. The variation of bandgap as a function of nitrogen composition at 300K has been found to be in good agreement with the predictions of the band anti-crossing model. An overall redshift for a composition of \( x=0.022 \) of 60meV has been found, corresponding to a reduction of \( \sim 27\)meV per percentage nitrogen. The temperature dependence of the PL peak energy for high nitrogen compositions also appears to exhibit an anomalous energy shift generally accepted to be a signature of dilute nitride alloys. The pacification of the effect for the highest composition sample (\( x=0.022 \)) is believed to be consistent with the predictions of the LCINS (Linear Combination of Isolated Nitrogen States) model. Additionally, photoluminescence and photomodulated reflectivity measurements on bulk InAs control samples indicate that, in contradiction to accepted understanding, the room temperature photoluminescence does not originate from a band-to-band transition.

Time-resolved pump probe measurements have been used to investigate the influence of doping level on spin lifetimes in InSb and InAs. In InSb a strong suppression of spin relaxation (\( \sim 30 \) times) with a degenerate doping level has been observed at 300K. The measured lifetime in a non-degenerate sample of \( \sim 16\)ps is in good agreement with theory, and indicates possibly comparable contributions of the D'yakonov-Perel and Elliott-Yafet mechanisms to spin relaxation. The insensitivity of that lifetime with respect to temperature is not consistent with theory, however, and is believed to indicate that, even at low doping levels, the Fermi level in InSb is close enough to the
Conclusions

band edge that the material exists in a regime intermediate between those described by the non-degenerate and degenerate formulations of the D’yakonov-Perel and Elliott-Yafet models. Furthermore, the degree of suppression in the degenerately-doped sample is larger than that predicted by the relevant theoretical expressions, indicating a possible error in their scattering mechanism dependent pre-factors. The mechanism responsible for the strong observed suppression is unclear at this time. In near-degenerately doped InAs a less significant suppression (>2 times) has been observed at 300K, with non-degenerate lifetime of ~20ps measured. In moderately n-type doped Hg₀.₇₈Cd₀.₂₂Te, the measured temperature dependence of the spin lifetime is in good agreement with that predicted by the Elliott-Yafet model. Therefore, it is believed the Elliot-Yafet mechanism is dominant in that material.

The spin galvanic effect and circular photo-galvanic effect have been studied using the FELIX free-electron laser in the Netherlands. Taking advantage of the facility’s ability to produce ultrafast pulses over a broadly tuneable wavelength range, it has been possible to perform the first measurements unambiguously identifying the spin galvanic effect, without the need for an applied magnetic field. Key to the success of these measurements, was the anisotropic nature of the linear spin-splitting in the presence of both bulk inversion asymmetry and structural inversion asymmetry in the GaAs quantum well structure studied. As predicted by a phenomenological model, along [110] suppression of the circular photo-galvanic effect current (which is proportional to the quantum well absorption) has been observed, permitting the identification of the dominant current in that case as due to the spin galvanic effect (which is proportional to the differential of the quantum well absorption). Furthermore, at photon energies close to the separation between the quantum well’s ground and first-excited states, the circular photo-galvanic effect current is expected to be negligible along all directions, so the non-zero current measured along [110] at such energies is a pure spin galvanic effect current.

Finally, high pressure measurements of InAs-based devices emitting at mid-infrared wavelengths (i.e. ~3.3μm) have been performed in an effort to study the efficiency of Auger recombination processes in type-I and type-II double-heterostructures. It was hoped that the type-II structure would serve to suppress Auger recombination. However, a strong increase in light output as a function of pressure in the type-II
device indicates the CHCC Auger process is the dominant current path (i.e. >74% of the total current). For the type-I device, radiative recombination and Auger recombination processes appear more comparable (i.e. <54% radiative contribution). The CHSH Auger process has been shown to be insignificant in both devices. The increased Auger recombination efficiency in the type-II device is attributed to the short tunnelling distances of carriers confined at the heterostructure interface, resulting in a small effective active region with a large carrier concentration. Therefore, although the Auger coefficient may be smaller in the type-II device as anticipated, the total Auger recombination rate (i.e. $Cn^3$) is in fact larger than that in the type-I device.
9 Appendix

9.1 The Varshni Equation

In a bulk semiconductor, the direct and indirect energy gaps are temperature dependent quantities. The most common function used to parameterise their dependence is the Varshni equation.[1] This is an empirical formula which defines the behaviour of the energy gaps in terms of three parameters as follows:[2]

\[ E_n(T) = E_n(T = 0) - \frac{\alpha T^2}{\beta + T} \]  

(9.1)

where \( T \) is the lattice temperature, \( E_n \) is the relevant energy gap at that temperature, \( E_n(T=0) \) is the energy gap at absolute zero, and \( \alpha \) and \( \beta \) are physically meaningless fitting parameters with units meV·K\(^{-1}\) and K respectively. Experimental values of \( E_n \) at low temperature are available for most common semiconductors, so \( E_n(T=0) \) is usually quite well-defined.

9.2 Interferogram Analysis

9.2.1 Introduction

In interferometry, a plot of fringe intensity as a function of mirror displacement (or retardation) is known as an ‘interferogram’. Encoded within an interferogram is the phase and intensity information required to construct a spectrum of the constituent
light. The interferogram and spectrum are related to one another by a Fourier transform (FT) pair as follows:[3]

\[ I(\Delta) = \int_{-\infty}^{+\infty} S(\Delta)e^{i2\pi\Delta\delta} d\Delta \] (9.2)

\[ S(\Delta) = \int_{-\infty}^{+\infty} I(\delta)e^{-(i2\pi\Delta\delta)} d\delta \] (9.3)

Where \( I(\delta) \) is the interferogram sampled as a function of retardation, \( \delta \) (in cm), and \( S(\Delta) \) is the spectrum sampled as a function of optical frequency (or wavenumber), \( \Delta \) (in cm\(^{-1}\)). Due to the units used in each case, it is convention to refer to the interferogram as being measured in the real space domain, while the spectrum is measured in the inverse space domain. It is equally valid to sample an interferogram in units of time delay, \( \tau \) (in s), corresponding to a spectrum with units of audio frequency, \( f \) (in s\(^{-1}\)).

A true (or ideal) interferogram is perfectly symmetric about ZPD (i.e. an even function). An asymmetric interferogram is, in fact, physically nonsensical. However, practical limitations such as sampling errors (i.e. not sampling symmetrically about, or at, ZPD), dispersion within the beamsplitter, and frequency dependent responses of any electronic filters\(^{20}\) used to process detector signals result in the addition of a phase error to a measured interferogram. This phase error manifests as imaginary (or odd) components in the measured interferogram, which is then re-defined as follows:[3]

\[ I(\delta) = \int_{-\infty}^{+\infty} S(\Delta)e^{i2\pi\Delta\delta} e^{i\theta(\Delta)} d\delta \] (9.4)

Where \( \theta(\Delta) \) is the phase angle, and is usually a smoothly varying function. The more component wavelengths are present in \( S(\Delta) \), the more rapidly their relative phases randomise (i.e. average to some constant value) as one moves away from ZPD.

\(^{20}\) Note this is not an issue in step-scanning FTIR systems as the modulation frequencies of all component wavelengths in the interferogram are identically zero.
Therefore, interferograms of broadband sources, such as an incandescent light bulb, are narrower than those associated with narrow band emitters such as lasers. In the former case, the interferogram signal typically becomes negligible (i.e. only a small fraction of the maximum signal) at retardations of the order of a millimetre, while for the latter it can be of the order of metres. The precise retardation at which the limit is reached is known as the coherence length of the source. This correlation between source bandwidth and coherence length is a fundamental property of FT pairs.

9.2.2 Sampling Frequency, Zero-Filling and Aliasing

In a practical instrument it is not possible to sample an interferogram continuously (i.e. with points which are infinitely close together). Rather, an interferogram is sampled in discrete intervals, leading to the re-definition of equations 9.2 and 9.3 as a pair of Discrete Fourier Transforms (DFT's):

$$I(n) = \frac{1}{M} \sum_{m=0}^{M-1} S(m) e^{\frac{2\pi inm}{M}}$$  \hspace{1cm} (9.5)

$$S(m) = \frac{1}{N} \sum_{n=0}^{N-1} I(n) e^{-\frac{2\pi inm}{N}}$$  \hspace{1cm} (9.6)

Where $m=0,1,2\ldots M-1$ and $n=0,1,2\ldots N-1$ index the points in the spectrum and interferogram respectively. Outside of those ranges, functions $S(m)$ and $I(n)$ are assumed to be periodic, so $S(m)=S(m+M)$ and $I(n)=I(n+N)$. Note, generally, $M$ need not equal $N$. However, in most commercial instruments FT calculations are performed utilising Fast Fourier Transform (FFT) algorithms, which dramatically reduce the number of operations necessary over a DFT. When applied, FFT routines usually require the data to be prepared in a very specific way, including that $M$ equals $N$ and that both are powers of 2 (i.e. 128, 256, 512 and so on). For $M,N=1024$ an FFT can speed up calculations by as much as 100 times, with that factor scaling exponentially with $M,N$.\[4\]

It may be noted that by adding additional null points to the beginning and end of an
interferogram, \( N \) can be artificially increased. This results in smaller sample spacing in the spectrum (i.e. a higher resolution). However, because these points are null, no information has been added to the measured spectrum by their introduction. Rather, the additional points result in an interpolation effect. Adding null points to an interferogram in this way is known as 'zero-filling', and is commonly used by FFT routines when transforming data sets with \( N \) not equal to a power of 2 (i.e. zeroes are added to make up the difference between \( N \) and the next highest power of 2).

The spacing between sampled points can be expressed in terms of the number of points per unit retardation, or sampling frequency, \( f_s \). If \( f_s \) is too low, an effect known as aliasing can result, whereby short wavelengths appear in a spectrum as longer wavelengths. For example, consider sampling a pure sinusoidal wave of wavelength \( \lambda \) in discrete intervals of retardation with \( f_s = \frac{1}{\lambda} \). As shown in Figure 9.1a, the sampled points do not uniquely define this wave, it being possible to draw another sinusoid of longer wavelength which also passes perfectly through the exact same points. Performing a DFT of the sampled pattern will, therefore, yield a Fourier component of the longer wavelength, in this case \( \frac{\lambda}{4} \), as well as the original shorter wavelength sinusoid. Aliasing is also illustrated for progressively larger \( f_s \) in Figure 9.1b and Figure 9.1c. There does exist a sample frequency threshold, above which it is not possible to draw any longer wavelength sinusoids through all the sample points. This threshold is shown in Figure 9.1d, where \( f_s = \frac{2}{\lambda} \).

![Figure 9.1](image-url)  
*Figure 9.1* Schematic showing the origin of aliasing for a sampled Fourier component of wavelength \( \lambda \) (red line). Only when the sampling frequency \( f_s \geq \frac{2}{\lambda} \) do the sample points (green dots) uniquely define the original wave; known as the Nyquist sampling limit for that Fourier component.
Appendix

An interferogram generally contains many Fourier component wavelengths, rather than just one. In order to prevent aliasing of any components in the spectrum, $f_s$ must be at least twice the largest wavenumber, $\Delta_{\text{max}} = 1/\lambda_{\text{min}}$, present in the interferogram. This is known as the Nyquist sampling limit:[5]

$$f_s \geq \frac{2}{\lambda_{\text{min}}} = 2\Delta_{\text{max}} \quad (9.7)$$

### 9.2.3 Sampling Limits and Apodisation

It is also practically impossible to sample an interferogram at infinite retardations. If the coherence length of the source is greater than the instruments maximum achievable retardation, there will appear a discontinuity between the last non-zero acquired point (i.e. the farthest sampled point from ZPD) and the unsampled region beyond. Sampling only to a finite retardation is mathematically equivalent to multiplying the true interferogram by a boxcar function, $D(\delta)$, of width equal to the maximum practical retardation, $\delta_{\text{max}}$. This is, in turn, equivalent to convolving the corresponding spectrum with the Fourier transform of the boxcar function; a sinc function, $2\delta_{\text{max}} \sin(2\pi \delta_{\text{max}})/(2\pi \delta_{\text{max}}) = 2\delta_{\text{max}} \text{sinc}(2\pi \delta_{\text{max}})$. The measured spectrum is, therefore, given by:[5]

$$S(\Delta) = \int_{-\infty}^{+\infty} D(\delta) I(\delta) e^{i2\pi \Delta \delta} d\delta \quad (9.8)$$

where $D(\delta) = 1$ if $-\delta_{\text{max}} < \delta < +\delta_{\text{max}}$, and $D(\delta) = 0$ if $\delta < -\delta_{\text{max}}$ or $\delta > +\delta_{\text{max}}$. If the spectrum in question consists of a single spectral line at frequency $\Delta_1$ (i.e. a single Fourier component interferogram), then the measured spectral lineshape is equal to the convolution of a delta function with a sinc function, yielding:[5]

$$S(\Delta) = \int_{-\infty}^{+\infty} 2\delta_{\text{max}} \text{sinc}(2\pi \Delta_{\text{max}}) S(\Delta - \delta) d\delta \quad (9.9)$$

The resulting sinc function-like lineshape is a property of a given instrument, and is
referred to as the instrument lineshape (ILS). Hence, for any true spectral line, the practical limitations of an instrument result in a broadening of the frequency distribution of that line, with decaying side-lobes extending on each side. The larger the interferogram signal at the point of maximum retardation, the greater the deviation of the measured ILS from the true lineshape.

The detrimental effect a finite interferogram sampling range has on the corresponding spectrum can be reduced by means of a mathematical processing technique known as ‘apodisation’. This involves multiplying the interferogram by a weighted window function (note not a boxcar function) prior to its FT. A number of choices for the form of such functions exist, although all are generally continuous between $0 \leq \delta \leq \delta_{\text{max}}$, and equal to 1 at ZPD and 0 at $\delta_{\text{max}}$. The key property of an apodisation function is, therefore, that it removes the discontinuity found at maximum retardation. Perhaps the simplest function fulfilling these criterion is the triangular apodisation function. Of course, the process of apodisation is equivalent to convolving the spectrum with the FT of the apodisation function, which ultimately alters the ILS. The choice of function is, therefore, a trade-off between the ILS signal amplitude, side-lobe amplitude, and width. Although a number of functions have been proposed (and are used), it is generally accepted that a triangular apodisation function gives a good balance between ILS properties.[5] A triangular apodisation function, $A_{\text{tri}}$, may be defined as:

$$A_{\text{tri}}(\delta) = 1 - \left| \frac{\delta}{\delta_{\text{max}}} \right| \quad (9.10)$$

**9.2.4 Phase Correction**

The process of phase correction entails the (mathematical) removal of the phase term, $e^{i\theta(\Delta)}$, in equation 9.4, where $\theta(\Delta)$ is generally not equal to $0, \pi, 2\pi$, etc (i.e. there are imaginary components present). There are three commonly employed methods of phase correction: the power transform, the Mertz method[6], and the Forman method[7].
Power Transform

This is a trivial operation which involves taking the magnitude of the measured spectrum as follows:

\[
|S(\Delta)e^{i\theta(\Delta)}| = \sqrt{S_{\text{real}}(\Delta)^2 + S_{\text{imag}}(\Delta)^2} \tag{9.11}
\]

where \(S_{\text{real}}(\Delta)\) and \(S_{\text{imag}}(\Delta)\) are the real and imaginary components of the complex spectrum respectively. The power transform is probably the simplest phase correction technique to apply. However, this method of phase correction has a non-linear effect on spectral noise, including the rectification of any noise which takes the spectral amplitude negative. This can lead to photometric error.[3,5] The rectification of negative spectral components also makes the method unsuitable for correcting interferometric PM signals; such features generally have negative-going lobes.

The Mertz Method

The Mertz method performs phase correction in the inverse space domain. Noting that \(S_{\text{real}}(\Delta) = S(\Delta)\cos(\theta(\Delta))\) and \(S_{\text{imag}}(\Delta) = S(\Delta)\sin(\theta(\Delta))\), the phase function \(\theta(\Delta)\) can be computed as follows:

\[
\theta(\Delta) = \tan^{-1}\left(\frac{S_{\text{real}}(\Delta)}{S_{\text{imag}}(\Delta)}\right) \tag{9.12}
\]

The phase corrected spectrum may then be calculated by multiplying equation 9.4 by filter function \(e^{-i\theta(\Delta)}\) (i.e. \(e^{i\theta(\Delta)}e^{-i\theta(\Delta)} = 1\)). If the phase function is smoothly varying, it can be estimated from a short double-sided interferogram. Time can thus be saved during acquisition by scanning the full retardation range on only one side of the interferogram (and only a short range on the other). The Mertz method also has a linear effect on noise, minimising photometric errors.[3,5] By obtaining the (instrument related) phase function from a spectrum known to be positive over the full spectral range (e.g. reflectivity, transmission, or system response measurements), this technique may also be used to correct PM signals acquired interferometrically (i.e. positive and negative lobes are reproduced accurately).
The Forman Method

The Forman method performs phase correction in the real space domain\textsuperscript{21}. The filter function $e^{-i\phi(\Delta)}$, where $\phi(\Delta)$ is calculated using equation 9.12 as for the Mertz method, is Fourier transformed to obtain a phase interferogram. That interferogram is then convolved with the measured interferogram to obtain the phase corrected interferogram. The Forman method possesses all of the advantages discussed for the Mertz method, and additionally permits digital filtering of the spectrum simultaneously with phase correction.\cite{note:forman} In the past this has proved useful in commercial instruments where memory and computation power were at a premium. However, in the era of modern desktop computing this advantage is somewhat academic. In addition to its being a more complicated algorithm to apply, if digital filtering is not required the Forman method is actually less efficient than the Mertz method, as the convolution operation requires more processor cycles to evaluate than multiplication.

9.3 Building the Step-Scanning FTPM System

9.3.1 The Step-Scan Interferometer

The step-scanning FTIR spectrometer employed for FTPM is based on the remnants of a retired Bomem DA3 series slow-scanning system. At the start of this project this consisted of the steel shell (suitable for evacuation), the mirror optics (Al spherical and flat), two sources (a quartz bulb with tungsten filament in halogen gas, and a graphite globar), a source pick-off mirror (aluminium 90° off-axis parabolic), and a quartz glass beamsplitter. However, both the moving and stationary retro-reflecting mirrors were missing, along with the opto-mechanical system which controlled the moving mirror (including the HeNe source/path used to gauge retardation in the as-bought configuration). There were also no electronics of any kind, so no means of remote control or data acquisition.

The system was originally designed to operate in a vertical stack, the upper portion containing all the critical components (e.g. sources, beamsplitter, retro-reflecting

\textsuperscript{21} Note the Forman method is mathematically equivalent to the Mertz method.
mirrors, etc.). The lower portion plays no part in the process of generating an interferogram. Initially intended as the instrument's sample chamber (i.e. the sample would have been placed inside, in the path of the collimated beam, for transmission-type measurements), it contains only three mirrors whose purpose is to direct the collimated beam through the sample, and focus it around ~4cm beyond the exit window. For this works application this segment of the instrument is entirely redundant. However, alignment of the upper portion is near impossible in any orientation other than the vertical (as it was originally designed to operate), which in turn makes extraction of the light from the instrument extremely difficult without the sample chamber. Thus the sample chamber is left in position. A schematic of this arrangement as it now appears, after modification, can be found in Figure 9.2.

The instrument height is around 2m, making it inherently rather unstable. The original stand/frame was not suitable for use in the designated lab, so a made-to-measure metal frame was manufactured which can be securely screwed to the optical table beneath. In addition to preventing lateral movement of the FTIR, the frame design also consisted of a spine which runs up the back-side of the instrument, with arms at around the level of the beamsplitter which come alongside and tightly grasp the front. This construction was intended to help reduce the degree of tilt or wobble in the system. Further bracing was provided by a metal bar across the top of the sample chamber (on its RHS as it appears in Figure 9.2), which when tightened forced the instrument into the table surface. Securing to the table in this way helped damp vibrations in the system, a critical problem in FTIR measurements where any movements on the order of optical wavelengths (~micrometres in the MIR) can destroy an interferogram signal (particularly close to ZPD where its differential is largest).
Figure 9.2 The step-scanning FTIR system as it appears today, post-refit. The source and alignment laser beam paths are shown with red and blue-dashed lines respectively. Both segments of the instrument may be sealed with covers. The sample chamber provides two alternative paths for light exiting the instrument, each at a different height above the table. The upper path is more convenient for coupling with any external FTPM optical arrangement.
Ideally the system would be evacuated for MIR applications, or at least purged with an inert gas such as dry nitrogen. Each segment is in fact vacuum sealable (in principle), and a valve on the instrument’s top provides a means of purging or evacuating (along with water cooling for the sources). However, successful evacuation of the instrument has proven difficult and, after several attempts to do so, efforts were eventually abandoned; vacuum sealing or purging the external FTPM arrangement was not convenient at this prototype stage either, and unless the entire FTPM system is similarly purged/evacuated atmospheric absorption would always affect spectra. Hence, evacuating/purging just one component at this point was not considered useful. However, all additional components were designed or purchased with evacuation/purging in mind, including using suitable electrical feedthroughs. It is hoped that in the future, now that proof of the FTPM principle has been successfully demonstrated, the issue of evacuation/purging can be revisited.

In order to restore the system to a working status, the first step was obviously to fit new moving and stationary retro-reflecting mirrors. Based on schematics provided by the author, the workshop at Surrey constructed extensions to house the new mirrors approximately where the system’s original mirrors were once positioned. A Newport VP-25XA precision translation stage was chosen as the base for the moving mirror. This stage was capable of minimum step-size of 0.1μm, with its position accurately measured relative to an internal glass scale. The housing for the moving mirror was, therefore, designed with the dimensions of this stage in mind. For the stationary mirror, a spring loaded adjustment was included to allow some tuning of ZPD. The mirrors themselves had gold first surfaces, and were of corner-cube design (each face ~2"x2"). These mirrors were chosen to facilitate alignment of the instrument, their main advantage over flat mirrors being that their angle of orientation is irrelevant with respect to the parallelism of the incident and reflected beam paths. In other words, any incident beam is always reflected back along a (displaced) parallel path. A mechanism for securely attaching the moving mirror to the translation stage was also machined by the workshop. The stage itself is controlled via a motion control unit, which in turn communicates with a lab PC via a standard GPIB interface.

---

22 This stage was rated as suitable for use in an evacuated or purged environment.
Appendix

With regard to the instruments beamsplitter, quartz is not suitable for MIR wavelengths beyond ~4µm. Since the intention was to measure beyond that limit, a KBr beamsplitter suitable for the original DA3 model (transmitting over the range ~2-22µm) was purchased from Bomem. However, the mounting for the beamsplitter had also deteriorated over time, so a new means of securing it into position was added. Nylon-tipped thumb screws are thus used to lock the beamsplitter into position.

Alignment of the spectrometer is performed using a HeNe laser whose path through the FTIR is the reciprocal of that followed by the light from the internal source(s). That is, the laser beam enters the instrument through the exit window and terminates at the source element (when appropriately aligned), as shown in Figure 9.2. Although KBr is essentially opaque at visible wavelengths, there is a small visible window around 1cm in diameter in the centre of the beamsplitter optimised for transmission at 632.8nm. In the original DA3 system, the HeNe calibration laser would have used this window. When the laser beam is expanded by a lens inserted before the beamsplitter (from the point of view of the laser), a circular fringe pattern may be observed at the position of the collimating mirror by inserting a white card. The beamsplitter can then be adjusted to optimise the fringe pattern; taken to be when the circular rings are seen to collapse towards (and expand from) a point at the centre of the pattern when the path difference is changed. Once that task is complete, the expanding lens can be removed, and the laser's path directed (using the two mirrors located after the beamsplitter), through the iris and onto the source element. The light emitted by the sources can then be expected to follow precisely the same path through the instrument (albeit in the opposite direction), and thus also give an optimised interferogram as a function of path difference. This beam path can be double-checked visually in the case of the tungsten-halogen (visible/NIR) source. If the aligning laser is also intimately 'tied' into the FTPM arrangement external to the FTIR, the optics in the FTIR instrument may be considered part of the FTPM setup itself. This aspect will be discussed further in the following section.

An acquisition program has been written in the graphical programming language LabVIEW 7.0, which both controls the moving mirror, and acquires the readings from various instruments (e.g. lock-in amplifier, multimeter, oscilloscope, etc.). At the end of each scan the interferogram is processed; it is apodised using a triangular window.
Appendix

function (see section 9.2.3), and phase corrected using the Mertz method (see section 9.2.4). Finally the interferogram is Fourier transformed using an FFT algorithm to obtain the spectrum. Averaging of multiple scans is performed in the spectral domain.

To locate ZPD, the mirror is scanned with a continuous (slow) velocity throughout its range (or near its expected location if an educated guess is possible), with the detector signal viewed as a dc level on an oscilloscope. At ZPD, a dramatic change in signal amplitude is observed, with the stage's position at that time displayed on the acquisition program's front panel. That position can subsequently be used as a point of reference for each scan. This location of ZPD is reasonably stable, remaining consistent (~±1μm) over several months' operation.

9.3.2 The FTPM Optical Arrangement

The fundamental principle underlying an optical arrangement in a PM experiment is, of course, largely independent of the spectrometer used to spectrally resolve the light. Hence, an FTPM set up has much in common with the general (PR) arrangement described in section 1.3.2. However, an optical layout has been devised for FTPM experiments that makes use of a single HeNe laser as a modulation source, an alignment beam for the step-scanning FTIR, and an alignment beam for the PM optical arrangement itself. In so doing, the FTIR itself essentially became part of the PM optical arrangement. The reason for taking this approach lies in the careful alignment of the spectrometer that is periodically required, particularly when the beamsplitter is changed, which can leave the beam exiting the spectrometer along a slightly different path and, hence, altering any external alignment also. As the FTIR and FTPM optical components are coupled in this way, it makes sense to couple their alignment accordingly.

A typical procedure for setting up an FTPM experiment may be summarised as follows. Using the HeNe laser, the external optics are first arranged (right up to the detector), using irises to provide alignment markers. Three flipper mirrors are required; one to select between modulation and alignment, the other two switching between alignment of the FTIR or PM optics. By design, if all three mirrors are out of the beam path, the arrangement is in the modulation (or experiment) mode.
Appendix

Obviously, the FTPM arrangement was built close to the exit window of the FTIR so that the spectrometer light could easily be coupled into the set up. It was important to ensure at this point that the both alignment paths (FTIR and PM) coincide at the positions the external alignment irises. This can be checked by alternately flipping the appropriate mirrors in and out of beam. Once convinced this has been achieved, one can begin optimising the FTIR alignment itself (as described in section 9.3.1). To raise the bench-top beams to the height of the spectrometer’s exit window, a periscope mirror arrangement is used. As close as can be managed, the laser beam should pass through the centre of the exit window, and hit the centre of the first mirror. If adequately aligned, the FTIR source light should then proceed cleanly through the FTPM optical arrangement and arrive at the detector position as intended.

Figure 9.3 shows the beam path for the experiment mode (i.e. all flipper mirrors out of beam) in an FTPT configuration. Figure 9.4 is a similar schematic of each alignment configuration. In both cases the optics internal to the FTIR is not shown; it has been illustrated previously in Figure 9.2. Every optical element outside of the FTIR is shown in both cases in Figure 9.4, though those irrelevant to each particular configuration are shaded out. The FTPR configuration is almost identical to that shown in Figures 9.3 and 9.4; only the optics from the sample onwards are different.
Figure 9.3 The FTPT optical arrangement external to the FTIR spectrometer for the modulation (or experiment) mode. Optical components not relevant to this mode are greyed out. The cryostat is optional, although in the MIR most PM experiments require it.
Figure 9.4 Alignment modes of the FTPT arrangement. The FTIR alignment optics are shown on the LHS, while on the RHS the PT arrangement optical alignment components are shown. Optical components not relevant to each mode are greyed out. Of particular note are the active flipper mirrors and the positions of each iris. All mirrors are, of course, adjustable. See Figure 9.2 for FTIR internal optics.
9.4 The Dielectric Constant and Basic Semiconductor Optical Properties

Maxwell's equations result in the following definition for the velocity of light, \( v \), in a medium:[8]

\[
v = \frac{c}{\sqrt{\varepsilon \mu}} = \frac{c}{n_r}
\]  
(9.13)

Where \( c \) is the velocity of light in free space, \( \varepsilon \) and \( \mu \) the dielectric constant and magnetic permeability of the medium respectively, and \( n_r \) is the complex refractive index. Semiconductors typically have magnetic permeability's close to unity, and equation 9.13 thus reduces to \( n_r = \varepsilon^{1/2} \).

The electric field component, \( E \), of an EM wave incident on a continuous dielectric material may induce polarisation, \( P \), defined as the electric dipole moment per unit volume. For an EM wave propagating in 1D (parallel to \( z \)) through a medium, \( E \) may be expressed incorporating the dielectric constant like so:[9]

\[
E = E_0 \exp \left[ i \left( \omega t - \varepsilon^{1/2} k_0 z \right) \right]
\]  
(9.14)

Where \( k_0 \) is the wavenumber in free space, and \( \omega \) is the wave's angular frequency. At this point one may note that generally the phase of the polarisation and the electromagnetic wave will not be the same. That is, in a real system the response of an electron to \( E \) is not instantaneous. This damping effect is accounted for by the imaginary part of the complex refractive index, (i.e. \( n_r = n - i \kappa \)). Substituting \( n_r \) for \( \varepsilon^{1/2} \) in equation 9.14, one obtains:

\[
E = E_0 \exp (- \kappa k_0 z) \exp \left[ i \left( \omega t - nk_0 z \right) \right]
\]  
(9.15)

Thus \( \kappa \) contributes an exponential decay to the propagating EM wave. For this reason \( \kappa \) is often referred to as the extinction coefficient, and is related to the absorption
coefficient (or intensity loss per unit path length), \( \alpha \), as follows:[8]

\[
\alpha = 2\kappa k_0 \quad (9.16)
\]

Noting then that \( \varepsilon^{1/2} = n - i\kappa \), the real and imaginary parts of a material's complex dielectric function \( (\varepsilon = \varepsilon_1 + i\varepsilon_2) \) may be written as follows:

\[
\varepsilon = n_r^2 = n^2 - \kappa^2 - i2n\kappa \quad (9.17)
\]

so

\[
\varepsilon_1 = n^2 - \kappa^2, \quad \varepsilon_2 = 2n\kappa \quad (9.18)
\]

The real and imaginary components of any complex variable obeying the laws of causality (i.e. effect under no circumstances may preceed cause) and linearity (i.e. if the cause is doubled so to is the effect) may be expressed in terms of one another via the Kramers-Kronig dispersion relations. This leads to the following expressions relating \( \varepsilon_1 \) and \( \varepsilon_2 \):[10]

\[
\varepsilon_1(\omega) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\omega' \varepsilon_2(\omega')}{\omega^2 - \omega'^2} d\omega' \quad (9.19)
\]

\[
\varepsilon_2(\omega') = -\frac{2\omega'}{\pi} P \int_0^\infty \frac{\omega \varepsilon_1(\omega)}{\omega^2 - \omega'^2} d\omega \quad (9.20)
\]

where \( P \) is related to the degree of polarisation within the material. These expressions thus allow one component of the dielectric constant to be determined based on knowledge of the other, although it should be noted their relation via integration over all frequencies implies this knowledge must be complete (i.e. known for all \( \omega \)) for an accurate reconstruction.

Considering only direct interband transitions between states in the conduction band and valence band of a typical semiconductor, \( \varepsilon_2 \) takes the following form:[10]
where $|M_{vc}|$ is the transition matrix element linking the conduction band state with energy $E_c$ and valence band state $v$ with energy $E_v$. A similar expression for $\varepsilon_1$ can be obtained by taking the Kramers-Kronig transform (equation 9.18) of equation 9.20. The surface integral, excluding $|M_{vc}|$, is the joint density of states, which expresses the occupation potential (or number of initial and final states) at energies $E_c$ and $E_v$ in the conduction and valence bands respectively. Hence, the larger the joint density of states, the more electrons (in the valence band) and vacant states (in the conduction band) might be available so that optical excitation can occur, and the greater the probability of a photon of energy $\hbar\omega=(E_c-E_v)$ being absorbed. It should be noted, therefore, that $\varepsilon_2$ is closely related to the absorption coefficient (compare also equations 9.16 and 9.18).

Inspection of equation 9.21 shows $\varepsilon_2$ is well-defined and smoothly varying for all finite photon energies, except when $|\nabla_k(E_c-E_v)|=0$ (i.e. where the joint density of states is infinite). At such points $\varepsilon_2$ exhibits a singularity, in the immediate vicinity of which it varies extremely rapidly. This condition is satisfied at the high symmetry points of the Brillouin zone, where the Bloch functions have (local) maxima or minima. Any optical characterisation technique sensitive to $\varepsilon_2$ may, therefore, be expected to have sharp resonances at photon energies corresponding to direct transitions at Brillouin zone critical points. The most important critical point in a typical (e.g. based on the zinc-blende lattices) direct gap semiconductor is the highly symmetric $\Gamma$-point (i.e. $(E_c-E_v)=E_g$), around which the density of states varies as $(E-E_g)^{1/2}$. This dependence is a consequence of the bands approximating a parabolic shape close to $E_g$, and is characteristic of a particular class of critical point; $M_0$ (or a minimum, corresponding to a minimum in the conduction band and a maximum in the valence band). It can be shown there are four distinct types of critical point which can occur within the first Brillouin zone of a typical bulk semiconductor; labelled $M_0$, $M_1$, $M_2$, and $M_3$. The energy dependences of the joint density of states around each critical point type are listed in Table 9.1 [10] It is the derivative of the joint density of states that is exploited
Appendix

(via its presence in $e_2$) by PM techniques when applied to semiconductors. Therefore, the constant $C$ in Table 9.1 is irrelevant in that context.

<table>
<thead>
<tr>
<th>Critical Point Type</th>
<th>Joint Density of States</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_0$ (minimum - CB min, VB max)</td>
<td>$0$</td>
</tr>
<tr>
<td>$M_1$ (saddle point - CB min, VB min)</td>
<td>$C - (E_0 - E)^{1/2}$</td>
</tr>
<tr>
<td>$M_2$ (saddle point - CB max, VB max)</td>
<td>$C$</td>
</tr>
<tr>
<td>$M_3$ (maximum - CB max, VB min)</td>
<td>$(E_0 - E)^{1/2}$</td>
</tr>
</tbody>
</table>

Table 9.1 The joint density of states in the vicinity of each of the critical point types found in a typical semiconductor’s Brillouin zone. CB and VB refer to conduction band and valence band respectively. C is a constant. The relevant constant of proportionality has been omitted.[10]

9.5 Spin-Half Particles in a Magnetic Field

An electron is a spin-half particle, or Fermion. The interaction of an electron’s spin with a magnetic field is of paramount importance in spintronics. The following is a basic quantum mechanical derivation of that interaction.[11]

A magnetic field, $B$, appears in an electron’s crystal Hamiltonian as an additional term $\mathcal{H}_S$, leading to the following form for time dependent Schrödinger equation (TDSE) describing its energy state:

$$ E \Psi(r,t) = \left[ \mathcal{H} + \mathcal{H}_S \right] \Psi(r,t) \quad (9.22) $$

where $\mathcal{H}$ is the Hamiltonian in the absence of the field (e.g. $\mathcal{H} = (\hbar k)^2 / 2m_e + V$), $E$ is the electron’s energy, and $\Psi(r,t)$ its wavefunction. Assuming the electron energy is at rest in free space, the equation 9.22 reduces to $E\Psi(r,t) = \mathcal{H}_S \Psi(r,t)$, and it follows from classical physics that the interaction energy (equivalent to $\mathcal{H}_S$) of the magnetic dipole moment due to an electron’s spin, $\mu$, and $B$ is given by:

$$ \mathcal{H}_S = -\mu \cdot B \quad (9.23) $$
where $\mu$ is related to the Pauli vector operator, $\sigma$, as follows:

$$
\mu = \left( -\frac{e\hbar}{4m_e} g \right) \sigma
$$

(9.24)

where $e$ is the electron charge, $m_e$ the electron mass, and $g$ is the electron spin g-factor (a number closely related to the gyromagnetic ratio; the ratio of a system's generated magnetic dipole moment to its angular momentum). It may be noted that $\sigma$ is directly related to the spin angular momentum operator, $S=\hbar/2\sigma$.

The matrix representations of any orthogonal Cartesian components of $S$ may are given by $S_x=\hbar/2\sigma_x$, $S_y=\hbar/2\sigma_y$ and $S_z=\hbar/2\sigma_z$, where $\sigma_x$, $\sigma_y$ and $\sigma_z$ are the Pauli matrices, which read as follows:

$$
\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
$$

(9.25)

These are in fact general representations of angular momentum for any two-dimensional basis. As the quantisation axis defined by $\sigma$ can be orientated along any direction within this coordinate system, it is useful to obtain a more general expression for an arbitrary component of $S$, $S_n$. To that end, one can define a unit vector, $\hat{n}$, with polar angles $\theta$ (vertical azimuth from positive $z$) and $\phi$ (horizontal azimuth from positive $x$), allowing one to write:

$$
S_n = S \cdot \hat{n} = \frac{\hbar}{2} \begin{pmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{pmatrix}
$$

(9.26)

where the Cartesian components of $\hat{n}$ have each been expressed in terms of those polar angles.

If $B$ lies parallel to the $z$-axis, $B=|B|\hat{z}$, where $\hat{z}$ is a unit vector pointing along the $z$-axis, combining equations 9.23 and 9.24 yields:
\[ \mathcal{H}_s = G|B|\sigma_z \quad (9.27) \]

where \( G = -(ehg/4m_e) \). The time independent Schrödinger equation (TISE), \( E\Psi = \mathcal{H}_s \Psi \), then reads:

\[ G|B|\sigma_z \chi = E \chi \quad (9.28) \]

where \( E \) are the energy eigenvalues of the permitted spin states, \( S_z = (h/2)\sigma_z \). The eigenvalues of \( S_z \) are \( m_s \hbar \), yielding:

\[ E_{m_s} = 2G|B|m_s, \quad \text{where } m_s = \pm 1/2 \quad (9.29) \]

The eigenvectors, \( \chi_{m_s} \), corresponding to these spin states, often labelled \( \alpha \) (for \( m_s = +1/2 \)) and \( \beta \) (for \( m_s = -1/2 \)), are general eigenvectors for a spin-1/2 particle representation (or basis vectors for two dimensional 'spin-space'):

\[
\alpha = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (9.30)
\]

with the systems eigenfunctions given by \( \chi = a\alpha + b\beta \), where \( a \) and \( b \) are constants representing spatial orientation with respect to the quantisation axis (i.e. if \( a = 1, b = 0 \) one has a spin-up eigenstate, while for \( a = 0, b = 1 \) one has a spin-down eigenstate).

Given these solutions of the TISE, one can deduce that the TDSE permits only two stationary solutions, \( \Psi_{m_s}(t) \):

\[
\Psi_{+1/2}(t) = \alpha e^{-i\omega t/2}, \quad \Psi_{-1/2}(t) = \beta e^{i\omega t/2} \quad (9.31)
\]

where \( \omega = E_{m_s}/m_s = 2G|B|/\hbar \). Therefore, the general solution of the TDSE may be expressed as a linear combination of these two spin states, yielding:

\[
\Psi(t) = c_1 \alpha e^{-i\omega t/2} + c_2 \beta e^{i\omega t/2} \quad (9.32)
\]
where \( c_1 \) and \( c_2 \) are constants. If \(|\Psi(t)|^2\) is normalised to unity (as is convention), \(|c_1|^2 + |c_2|^2 = 1\). By determining the eigenfunction of \( S_n \), \( \Psi_n \), at \( t=0 \) (using equation 9.26) it possible to find \( c_1 \) and \( c_2 \):

\[
\Psi_n(t = 0) = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} \end{pmatrix}
\] (9.34)

Thus, for spin-up injection along the z-axis (i.e. \( S_n \parallel B \)) one has \( \theta=0^\circ \) (the angle \( \phi \) being irrelevant in this orientation), yielding:

\[
\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\] (9.35)

The expectation value for a measurement made along any arbitrary direction within the established Cartesian frame, \( \langle S_n \rangle \), where \( S_n' \) need not be the same as \( S_n \), is given by:

\[
\langle S_n' \rangle = \int \Psi_n(t)^* S_n' \Psi_n(t) \, d\eta
= \langle \Psi_n(t) | S_n' | \Psi_n(t) \rangle
\] (9.36) (9.37)

where the second expression is the same as the first only expressed in Dirac notation.

The expectation value itself tells one the average result of a large number of similar measurements performed on identically prepared (and hypothetical) systems. Thus, although an individual measurement of spin must always yield \( \pm \hbar/2 \), the expectation value informs of the probability of finding each result. For example, the result \( \langle S_n \rangle = 0 \) implies that over a large number measurements the two outcomes \( \pm \hbar/2 \) are obtained with equal probability (i.e there is a 50:50 chance of either). Such a spin is then said to
exist as a mixture of both spin states, simultaneously exhibiting the character of both in equal parts. This is known as the superposition principle, in accordance with the ‘Copenhagen Interpretation’ of quantum mechanics due to Niels Bohr.

With this interpretation in mind, one can consider measuring the component of $\mathbf{S}$ parallel to the $z$-axis, the system having been initialised as in equation 9.35 previously (i.e. $S_n\parallel z$):

$$\langle S_z \rangle = \left< c_1, a e^{-i\alpha} \right| S_z \right| c_1, a e^{-i\alpha} \right> = \frac{\hbar}{2} \quad (9.38)$$

Thus, a spin injected parallel to $\mathbf{B}$ will remain in that state indefinitely unless acted upon by an external influence. It does not evolve with time. It can be shown in a similar way that $S_x$ and $S_y$ are identically zero at all times. This particular arrangement (i.e. $S_n\parallel B$) corresponds, therefore, to one of the stationary solutions of the TDSE (see equation 9.31), the other found when the spin is initialised ‘down’ parallel to the $z$-axis (i.e. $S_n\text{anti}\parallel B$). These stationary states are often referred to as the system’s ‘spin-eigenstates’.

Consider now the case of spin-up injection parallel to the $x$-axis. Equation 9.35 then becomes:

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \sqrt{2} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \quad (9.39)$$

and again measuring parallel to the $z$-axis one obtains:

$$\langle S_z \rangle = \left< c_1, \alpha^T e^{-i\alpha} + c_2, \beta^T e^{-i\alpha} \right| S_z \right| c_1, \alpha e^{-i\alpha} + c_2, \beta e^{-i\alpha} \right> = 0 \quad (9.40)$$

23 In this case one can deduce this result from $\langle S_z \rangle$ as $\langle S_z^2 \rangle = \langle S_x^2 \rangle + \langle S_y^2 \rangle + \langle S_z^2 \rangle = (\hbar/2)^2$ must hold true.
where the orthonormality relations $\alpha^T \alpha = \beta^T \beta = 1$ and $\alpha \beta = \alpha^T \beta = \alpha \beta^T = 0$ have been used. This result indicates that, as measured along the z-axis, the spin is in a superposition of both $\pm \hbar/2$ spin-states, with an equal probability of being found in either (as described previously). A similar calculation for $\langle S_x \rangle$ and $\langle S_y \rangle$ yields:

$$\langle S_x \rangle = \frac{\hbar}{2} \cos \omega t, \quad \langle S_y \rangle = \frac{\hbar}{2} \sin \omega t \quad (9.41)$$

Hence, the expectation value of $\langle S \rangle$ appears to precess about the z-axis in the $xy$-plane (at $z=0$) with angular frequency $\omega = 2G |B|/\hbar$. This situation is illustrated in Figure 9.5a, where a plot of $\langle S_x \rangle$, $\langle S_y \rangle$ and $\langle S_z \rangle$ as a function of time is also shown. Thus a measurement of $S_x$ will return a pure spin-state (i.e. $\langle S_x \rangle = \pm \hbar/2$) when $t = n\pi/\omega$, where $n=1,2,3,$ etc. Similarly for $S_y$, pure spin states are found when $t = (n+1/2)\pi/\omega$, where this time $n=0,1,2,$ etc.

The result of a similar analysis for $S_n$ not parallel to any of the three Cartesian axes is shown in Figure 9.5b. In this case $\theta = 55^\circ$ and $\phi = 0^\circ$. The expectation value $\langle S \rangle$ again precesses about the z-axis as in the previous example. However, the rotation plane is now vertically shifted such that $\langle S_z \rangle \neq 0$. Rather, $\langle S_z \rangle$ takes a constant positive value (i.e. $0 < \langle S_z \rangle < \hbar/2$), indicating there is a greater probability of obtaining the spin-up result in a measurement than spin-down, independent of time. The elevation of $\langle S \rangle$ with respect to the $xy$-plane in this configuration prevents the $S_x$ and $S_y$ components from ever achieving a pure spin-state, with the radius of the circle traced out by $\langle S \rangle$ as it precesses being smaller than $\hbar/2$. Of more academic interest is the plot of $\langle S_n \rangle || \langle S_n \rangle$ in Figure 9.5b, which exhibits a time-dependent behaviour somewhere between that of $\langle S_x \rangle$ and $\langle S_y \rangle$.

To summarise, when placed in a uniform and constant magnetic field, $B$, the expectation value, $\langle S \rangle$, of a spin injected parallel to any arbitrary direction within an observer defined Cartesian coordinate system precesses about the field axis with angular frequency $\omega = 2G |B|/\hbar$. Only if the spin is injected parallel to $B$ (i.e. $S_n || B$) is that precession eliminated, corresponding to a stationary solution of the TDSE, or spin-eigenstate. Only in such a state will a measurement of the spin component
parallel to \( B \) (i.e. \( S_n \parallel B \)) yield only one of \( \pm h/2 \) with certainty (i.e. \( \langle S_z \rangle = \pm h/2 \)). For all other orientations of spin the probabilities of measuring either value at any given time are non-zero (and constant), and the electron is said to be in a superposition of the two states (i.e. \( -h/2 < \langle S_z \rangle < +h/2 \)). A similar measurement of such a spin in the plane perpendicular to \( B \) (i.e. \( S_n \parallel S_x \) or \( S_n \parallel S_y \)) yields an expectation value which varies sinusoidally between \( +h/2 \) (or pure spin-up) and \( -h/2 \) (or pure spin-down). It is worth noting that these behaviours are the quantum mechanical analogues of the classical Larmor precession of a magnetic dipole in a uniform and constant magnetic field.

Figure 9.5 Precussion of the expectation value of a spin, \( S_z \) in a uniform and constant magnetic field, \( B \| z \), when injected a) along \( x \) and b) along an arbitrary direction (\( \theta = 55^\circ \)). The LHS shows a vector diagram illustrating the behaviour while the RHS shows corresponding plots of \( \langle S_x \rangle \), \( \langle S_y \rangle \), \( \langle S_z \rangle \) and \( \langle S_n \rangle \) (only in b), as in a) \( \langle S_n \rangle = \langle S_z \rangle = \langle S_x \rangle \).
9.6 References


