Efficiency Limiting Processes and Optimisation of Silicon Compatible Lasers for Optoelectronic integration and Optical Interconnects

by

Graham William Read

Submitted for the degree of Doctor of Philosophy

Advanced Technology Institute and Department of Physics
Faculty of Engineering and Physics Sciences
University of Surrey

August 2015
Abstract

Optoelectronic integration on silicon is an area of increasing interest for both physicists and the microelectronics industry. Due to the limitations of silicon as an optical gain medium, the integration of III-Vs with silicon microelectronics has become a prominent area of research. However, the fundamental physical differences between these materials has caused such lasers to be strongly limited by non-radiative recombination. Studies of these mechanisms are therefore essential for solutions to be developed that will allow commercially viable III-V/Si lasers to be fabricated. This thesis presents such studies for three of the four leading approaches to producing III-V/Si lasers (quantum dots on silicon are not studied), with conclusions on the relative performance of each presented in the final chapter.

AlGaInAs/InP laser active regions wafer bonded onto pre-processed silicon-on-insulator waveguides have exhibited strong performance, with electrical injection lasing demonstrated at room temperature. However, large and temperature sensitive threshold current densities of $\sim 3.4 - 6.16 \text{ kAcm}^{-2}$ indicate that the devices are not yet optimised. Defect current fractions of 22 - 39 % suggest that significant densities of threading dislocations propagate to the active region during bonding. In addition low $T_0$ and $T_1$ values, a lack of carrier density pinning and Z parameter values in excess of three suggest the presence of both carrier leakage and inter valence band absorption.

Development of the GaNAsP active material, has allowed lattice matched, direct epitaxial growth on silicon. The polar/non-polar III-V/Si interface and thermal expansion coefficient mismatch however, cause large densities of defects to form. As such, up to 68% of carriers are found to recombine non-radiatively via defect states. Improvements to performance are achieved by the use of MQW structures and by optimising the silicon surface orientation to minimise the formation of anti-phase domains, leading to an 18% increase in radiative current fraction. However, defect densities remain large, with additional thermally activated defects potentially caused by the diffusion of nitrogen from the QW, forming defect states at the QW/barrier interface. These states may also form the carrier leakage path, responsible for up to 27% of recombination.

Optimisation of the GaNAsP lasers by optical simulation predicts a potential increase in optical confinement factor from 0.35% to 0.6 % and 1.33% to 1.73%, corresponding to reductions in threshold current of 41% and 27% for single and multiple quantum well structures, respectively. Poor electrical performance was investigated by SEM of the contacts. This also identified limitations to the lithography, etching and metalisation, which caused among other effects, the burning of contacts under electrical injection. A processing optimisation study eliminated the contact burning, improved the IV characteristics and increased the facet output power by almost an order of magnitude. The addition of a post metalisation annealing step was also found to reduce the p and n-type contact resistances by 64% and 20% respectively.

A final method studied in this thesis is the use of GaInSb composite quantum wells grown on GaSb. III-SbS have been demonstrated to accommodate significant strain as well as tending to prevent dislocations from propagating in the growth direction. Lasing at room temperature with a threshold current density of $426 \text{ Acm}^{-2}$ is observed, with a radiative current fraction of 41%. Carrier leakage is found to be the dominant source of loss, accounting for 39% of recombination. However, based on pressure dependence studies an increase in carrier confinement of only 20 meV may be enough to halve this. The remaining 20% of carriers are thought to recombine via Auger processes, particularly CHSH due to the small difference between the bandgap and the spin-orbit splitting.
Acknowledgments

I must firstly thank Prof. Stephen Sweeney for his guidance and shared knowledge in acting as my supervisor. His consistent efforts in this role opened up the opportunities which allowed this thesis to be completed. I would like to give additional thanks for his efforts beyond those required for his role as my supervisor. His unwavering commitment and support to ensuring that I achieved my goals, particularly in the most difficult times of this process, is something I am immensely grateful for.

I would also like to thank Dr. Igor Marko for his extensive expertise and assistance in experimental work, particularly with the pressure system. Also Dr. Nadir Hossain, from whose work I expanded on in this thesis for his efforts as a mentor to me. Thanks also go to Dr. Shirong Jin and Dr. Kostis Michelakis for their help with simulations and device processing respectively. Dr. Michael Lewis, Dr. Barnabus Prouse and Dr Ben Crutchley I owe thanks for the hours of discussion, but particularly their friendship. The support staff, Chris Buxey, Tony Corless and John Underwood I appreciate greatly for keeping the labs and equipment running, advising me in their correct usage, and fixing them when that advice was ignored.

From the Philipps-University, Marburg I would like to thank Dr. Martin Zimprich, Dr. Bernardette Kunert, Prof. Kerstin Volz and Prof. Wolfgang Stolz for the Ga(NAsP)-based lasers and materials, as well as the interesting and enjoyable time I spent working within their group. To Prof. John Bowers at the University of California, Santa Barbara for the AlGaInAs silicon-evanescent lasers and Prof. Eric Tourniè at the Université de Montpellier for the GaSb based lasers, the studies of which I found fascinating.

The University of Surrey (UniS), Advanced Technology Institute (ATI) and Engineering and Physical Sciences Research Council (EPSRC) should also be acknowledged for providing the facilities and funding required for this PhD.

My parents, I cannot find the words to thank enough for their limitless love, enthusiasm and support throughout my life. Everything I have achieved I owe to them in some way. Thanks also to Iain for the countless hours of laughter that kept me me positive, sorry for not writing this as elegantly as you would have.

Lastly to Hannah, thank you for your infinite patience with me, I would not be writing this without you. I couldn’t have imagined a better partner, your love and commitment is something I look forward to spending the rest of my life trying to repay you for.
Publications


Conference presentations


Contents

1 Introduction ......................................................... 1
  1.1 Motivation ....................................................... 1
  1.2 Optoelectronic integration on silicon ......................... 3
    1.2.1 Silicon based lasers ....................................... 4
    1.2.2 III-V integration on silicon ......................... 5
      1.2.2.1 Hybrid integration .................................... 6
      1.2.2.2 Epitaxy of III-V active regions on silicon ............ 7
  1.3 Thesis aims ....................................................... 9
  1.4 Thesis outline ................................................. 10

2 Theory ................................................................. 12
  2.1 Semiconductor fundamentals .................................. 12
    2.1.1 Electronic band structure ................................ 12
    2.1.2 Bulk density of states .................................... 14
    2.1.3 The Fermi-Dirac distribution ............................... 14
    2.1.4 Carrier densities ....................................... 15
  2.2 Radiative electronic transitions ............................... 16
    2.2.1 Absorption ............................................... 16
    2.2.2 Spontaneous emission .................................... 17
    2.2.3 Stimulated emission ...................................... 17
  2.3 Non-radiative carrier recombination .......................... 18
    2.3.1 Defect-related recombination .............................. 18
    2.3.2 Auger recombination ..................................... 19
    2.3.3 Carrier leakage ........................................ 22
  2.4 The pn junction ............................................... 23
  2.5 Gain and population inversion .................................. 25
    2.5.1 Population inversion ...................................... 25
    2.5.2 Gain ..................................................... 27
  2.6 The cavity and achieving lasing threshold .................... 28
    2.6.1 The Fabry-Perot laser cavity ............................... 28
    2.6.2 The distributed feed back cavity (DFB) .................... 30
    2.6.3 Threshold gain and threshold carrier density ............. 31
    2.6.4 Free carrier absorption (FCA) and inter-valance band absorption (IVBA) .......................... 33
  2.7 Threshold current and slope efficiency ........................ 34
  2.8 Internal quantum efficiency .................................... 35
  2.9 Z analysis ...................................................... 36
2.9.1 Calculating Z ................................................. 36
2.9.2 Determining recombination mechanism fractions through Z analysis ............ 37
2.10 Characteristic temperatures, $T_0$ and $T_1$ ................................................. 38
2.11 Hydrostatic pressure .................................................................................. 39
  2.11.1 Influence on the crystal and band structure ........................................... 39
  2.11.2 Semiconductor laser characterisation as a function of hydrostatic pressure ... 40
2.12 Performance improvements ........................................................................ 42
  2.12.1 Quantum confinement ........................................................................... 42
  2.12.2 Optical confinement ............................................................................. 44
2.13 Bandgap renormalisation ............................................................................ 45

3 Experimental Processes .................................................................................. 46
  3.1 The general electroluminescence characterisation arrangement ...................... 46
  3.2 Device mountings ....................................................................................... 48
    3.2.1 The probe station ................................................................................ 48
    3.2.2 The single contact laser clip ............................................................... 49
    3.2.3 The two pin laser mount ................................................................. 49
  3.3 The closed cycle cryostat ........................................................................... 50
  3.4 High hydrostatic pressure system ................................................................ 51
  3.5 Device processing ...................................................................................... 54
    3.5.1 Lithography ......................................................................................... 54
    3.5.2 Etching ................................................................................................ 55
    3.5.3 Metalisation ......................................................................................... 56
      3.5.3.1 The Edwards E306A evaporator ................................................... 56
      3.5.3.2 The JLS loadlocked sputterer ....................................................... 56
      3.5.3.3 The PPC RTM 2016-M-2G-FC (208) rapid thermal anneal furnace ... 57

4 Efficiency Limiting Processes of Hybrid AlGaInAs Silicon-Evanescent Lasers .... 58
  4.1 Introduction and background ........................................................................ 58
    4.1.1 Hybrid lasers: an approach to optical integration on silicon microelectronics . 58
    4.1.2 Structure, composition and fabrication of hybrid AlGaInAs silicon-evanescent lasers ... 59
  4.2 Distributed feed back AlGaInAs-silicon evanescent laser characterisation ........ 63
    4.2.1 Electroluminescence .......................................................................... 63
    4.2.2 Spontaneous emission ....................................................................... 67
    4.2.3 Characteristic temperature ................................................................ 70
    4.2.4 Z analysis ......................................................................................... 72
    4.2.5 Spectral analysis ............................................................................... 78
    4.2.6 Distributed feed back AlGaInAs-silicon evanescent laser summary and conclusions ... 88
4.3 Fabry-Perot cavity hybrid AlGaInAs-silicon evanescent laser .................................................. 89
  4.3.1 Ambient temperature characterisation .................................................................................. 90
  4.3.2 Fabry-Perot cavity AlGaInAs-silicon evanescent laser low temperature characterisation 97
4.4 Conclusions ............................................................................................................................. 110
  4.4.1 Suggestions for device improvement .................................................................................. 111

5 Development and Optimisation of Monolithic GaNAsP/Si Quantum Well Lasers 113
  5.1 Introduction and Background ................................................................................................. 113
    5.1.1 Monolithic integration of III-Vs on Silicon as a solution to optoelectronic integration 113
    5.1.2 The band anti-crossing model for dilute nitrides ......................................................... 115
    5.1.3 Previous characterisation of the GaNAsP material system ........................................ 118
  5.2 The dependence of efficiency limiting processes on GaNAsP/Si quantum well number ..... 123
    5.2.1 Temperature dependent electroluminescence ............................................................... 125
    5.2.2 Hydrostatic pressure dependent electroluminescence .................................................. 136
    5.2.3 Conclusions ..................................................................................................................... 140
  5.3 The effect of silicon surface orientation on GaNAsP/GaP/Si QW laser non-radiative recombi-
    nation ....................................................................................................................................... 142
    5.3.1 Defect generation at the polar-non-polar GaP/Si interface ........................................... 142
    5.3.1.1 Anti phase domains (APDs) and anti phase boundaries (APBs) ............................. 143
    5.3.1.2 Optimisation of the GaP buffer for defect free III-V active material growth 
on Si ................................................................................................................................. 145
    5.3.2 Temperature dependent electroluminescence ............................................................... 145
    5.3.3 Pressure dependent electroluminescence ...................................................................... 149
    5.3.4 Conclusions ..................................................................................................................... 151

6 Simulation of the GaNAsP/BGaP/Si material system 153
  6.1 Optical simulation of single and multi-quantum well GaNAsP/Si laser structures ............ 153
    6.1.1 Waveguiding theory ....................................................................................................... 154
    6.1.2 Simulation details and methodology .............................................................................. 156
    6.1.3 Single quantum well structure optimisation .................................................................. 159
    6.1.4 Multiple quantum well structure optimisation ............................................................ 167
    6.1.5 Optical simulation conclusions ..................................................................................... 173
  6.2 GaNAsP/BGaP/Si wafer processing procedure development and device characterisation 
    ................................................................................................................................................. 176
    6.2.1 Identification of the processing limitations of post structure #16195 devices through 
    optical and scanning electron microscopy ........................................................................... 176
    6.2.2 Wafer-processing optimisation investigation ............................................................... 181
      6.2.2.1 Lithography ............................................................................................................. 183
      6.2.2.2 Etchant investigation .............................................................................................. 184

vii
Chapter 1

1 Introduction

1.1 Motivation

The demand for both data processing power as well as internet access and versatile online content are growing exponentially, from both developed and increasingly developing nations [33]. The result is a substantial increase in energy consumption attributed to computing, data centres and communications. It is predicted that at current growth rates the energy consumption associated with the internet alone could theoretically reach between 22% and 100% of the global electricity supply by 2023 (see figure 1), assuming annual reductions in power consumption per $Gbs^{-1}$ of throughput of 15% (optimistic) or 0% (pessimistic), respectively [160]. Since demand will not decrease, the efficiency and flexibility of the hardware and systems at the heart of the network infrastructure must be increased.

![Figure 1.1: The trends in power consumption of various aspects of internet infrastructure over time compared to Global electricity supply [160].](image)

Modern microelectronics are based almost exclusively on the semiconductor material silicon and the associated complementary metal oxide semiconductor (CMOS) technology. It was initially predicted by Gordon Moore that the number of transistors per unit area would double every two years [106], but has been generalised since to predict a doubling of processing power every two years. Transistors have been consistently shrinking (with 7 nm being the current smallest transistor base) in the pursuit of ever greater densities, which therefore are tending to a finite limit, but are also facing issues such as increased gate delay, resistances and power density.

Novel approaches to chip design such as tri-gate transistors [32] and high-k materials [55] are now required to continue to improve chip performance. Thus in an attempt to continue to meet the growth predicted by
Moore’s law, microprocessing architecture has moved towards parallel processing systems that consist of multiple individual, and increasingly specialised processing units operating simultaneously in parallel. This leads to an issue beyond that of energy consumption, in that demand for bandwidth across servers, boards and on-chip is also increasing rapidly to keep pace with ever faster parallel processors.

Electrical data transfer reaches its practical limit at a bandwidth distance product of $\sim 100 \text{ Gbs}^{-1}m$ \cite{154} due to the fundamental physical limitations of electrical interconnects, imparting noise and loss limits that prevent greater throughput being achieved \cite{93}. Increasing the interconnect dimensions will cause either the resistance or capacitance to rise, leading to a larger RC constant and increasing the number of I/O (input and output) pins would lead to larger and therefore more costly chips.

Optical data transfer has proven to be superior to electrical methods over larger distance scales, leading to its adoption for communications ranging from 1000s of km such as in the transatlantic communications cable down to a few metres between server racks in local area networks (LAN). Optical signals exhibit large bandwidth, low attenuation, power consumption and heat dissipation as well as eliminating cross-talk between interconnects in close proximity. With cooling and data transfer accounting for the bulk of data centre energy consumption, the overall energy demands of a photonic system would be vastly reduced \cite{46}. Optical data transfer systems generally operate at either the zero dispersion or minimum loss wavelength of silica fibres and Si/SiO$_2$ waveguides, depending on which process is limiting in a given situation. Finally, the already intrinsically high single mode bandwidth can be increased further by processes such as wavelength division multiplexing (WDM). In this process 10s to 100s of signals are passed down the same waveguide with slightly different wavelengths allowing them to be separated on arrival without interference being an issue. This is of particular benefit for the ever expanding adoption of parallel processing.

Based on the above it would be logical to reduce the scale of optical interconnects down sufficiently for inter- and eventually intra-chip communications, monolithically integrated within a Si substrate on which the microelectronic components are defined. In doing so such a system would exhibit the benefits of both technologies. Data processing using CMOS microelectronics provides abundant and low cost raw materials, high attainable material quality and mature processing standards and systems, whilst an optical data transfer system’s benefits include increased SNR (signal to noise ratio), increased bandwidth and low losses and heat dissipation. The main components required for such a system have mostly already been realised in a viable form, including modulators operating up to 50 Gbs$^{-1}$ \cite{47}, Ge/Si photodetectors, filters, waveguides, switches and even micro-electro-mechanical systems, enabling dynamic control of optical signals \cite{156}. The last essential component yet to be sufficiently developed for commercialisation is a suitable light source on silicon, specifically a laser.

The problem with a silicon based laser is its indirect bandgap, which causes gain within bulk Si to be inherently weak due to inefficient radiative recombination, with $\sim$1 photon produced per $10^6$ electrons injected \cite{122} and non-radiative recombination mechanisms prevalent \cite{167}. The difference between a direct and indirect bandgap is shown in figure 1.2. The minimum of the conduction band in a direct bandgap material is found at the same wavevector value as the valence band maximum, whereas in an indirect material
it is not. Photons convey little momentum (\(\hbar k\)) and so radiative recombination of two carriers must take place between states of near identical wavevector, which may take place readily in a direct bandgap material (1.1 left).

If the material has an indirect bandgap this cannot take place since momentum cannot be conserved. In order for the transition to take place a third body of sufficient momentum (a phonon) is required to interact with the carriers. Thus, the rate of radiative recombination is dependent on the density of not only carriers within the correct states but also phonons of sufficient momentum. The average lifetime of phonon assisted radiative recombination is of the order of \(\sim 250 \, \mu s\) compared to only \(\sim 10 \, \text{ns}\) for Auger, Shockley-Read-Hall (SRH) and free-carrier absorption leading to these process dominating and the internal quantum efficiency being exceptionally low [151].

![Spontaneous emission](image1)

![Wavevector (k)](image2)

Figure 1.2: Left: Direct bandgap dispersion relation. Right: Indirect bandgap dispersion relation.

One alternative is to integrate established III-V semiconductor laser material systems onto silicon microelectronics and make use of the beneficial properties of each. This method has its own inherent issues though. The lasers, detectors etc used for LAN communications are discrete components which limit both the size and efficiency of the system into which they are integrated to such a degree that they are not viable for small scale applications. Monolithically integrating III-V materials onto silicon by epitaxy is also problematic. The large lattice constant and thermal expansion coefficient mismatch between common III-V materials and Si, in addition to the fact that a polar/non-polar bond interface is formed [40]. These factors induce strain in the epitaxially deposited layers, which when sufficiently large, is relaxed by the formation of large densities of crystal defects during growth, rendering both efficiency and lifetime impractically low.

### 1.2 Optoelectronic integration on silicon

In order for a laser to be integrated with silicon based microelectronics, forming an optoelectronic integrated circuit (OEIC), it must overcome the issues discussed above. Further to this, the implementation of the laser
on Si must be compatible with existing CMOS processing procedures and infrastructure for the complete OEIC to be commercially viable and competitive with the non-photonic contemporaries. CMOS process compatibility dictates that any further processes carried out have no impact on the structures and components already defined within the wafer and are achievable without requiring major changes to the fabrication infrastructure. Several of the key requirements (taken from [51]) are as follows:

- A temperature of 450 °C must not be exceeded to prevent alloying of metals.
- The thermal-time budget for all elevated temperature (above ambient) processes should be minimised to prevent diffusion effects.
- Electrical discharge or introduction of mobile ions into the CMOS structure is to be avoided.
- Mechanical stresses or electrical perturbations such as parasitic capacitances, resistive leakage paths or shorts to the ICs must not be introduced.
- The maximum miscut angle of the Si substrate may not exceed 0.5 ° from the [001] direction [83].

Several approaches to achieving a silicon compatible laser have been investigated, which fall into two main categories. Fully silicon based lasers that attempt to directly address the flaws of Si as an active optical material or III-V-on-Si integration methods, which combine mature III-V active materials with Si substrates in such a way as to circumvent the lattice and thermal expansion coefficient mismatches.

### 1.2.1 Silicon based lasers

Many varied and often complex approaches to improving the light emission properties of silicon have been investigated in the hope of achieving an all silicon laser. The obvious benefit of such an approach is that CMOS compatibility should be more easily achieved and that the complexities introduced by combining dissimilar materials with Si are eliminated. One of the categories of silicon based laser is so-called low dimensional silicon, this encompasses all approaches that make use of structures that strongly confine carriers spatially, the most extreme case of which being a quantum dot (QD), which confines the electrons in all three spatial dimensions. The impact of LD-Si structures on the carriers is analogous to the “particle in a box” situation often used as an example of electron confinement in quantum mechanics.

As electrons are increasingly confined, both in scale and number of dimensions the allowed energies of the electrons and holes both increase (widening the effective bandgap) and become more discretised, tending towards the properties of an atomic electron rather than one which is found within a crystal lattice. In addition, since Heisenberg’s uncertainty principle states $\Delta x \Delta p \geq \hbar/2$ spatial confinement (minimising $\Delta x$) necessarily increases the uncertainty in momentum $(\Delta p = \hbar \Delta k)$. Therefore increased confinement leads to a greater spread of electrons in k-space, reducing the extent to which phonons are required for radiative recombination to be achieved and thus increasing the internal quantum efficiency of indirect gap materials such as Si [151]. Low dimensional Si has been leveraged to achieve optically pumped lasing in nanocrystals [91], porous Si [30], Si nano-pillars [3] and Si-insulator superlattices.
Stimulated Raman scattering (SRS) has also been used to achieve continuous wave (CW) optically pumped net gain at room temperature [26]. Scattering of photons by particles smaller than the wavelength (ions for example) is usually an elastic process known as Rayleigh scattering, electrons are absorbed and the re-emitted in a random direction with the same energy. However, with a probability of \( \sim 1 \times 10^{-7} \) the photons may scatter inelastically, either from an excited state, falling to a ground state and thus increasing the photon energy or from a ground state and relaxing to an excited state and losing energy.

These processes are collectively known as Raman scattering with the gain in energy known as the Stokes shift and the loss of photon energy known as the anti-Stokes shift. Although this process is very weak, if the Stokes photons are selectively confined or intentionally injected (signal beam) in addition to a pump light source then the Raman scattering process can be “stimulated”. The rate of Raman scattering is increased beyond its “spontaneous” rate as a function of Stokes photons and can effectively amplify these photons, generating net gain [25].

Si-Ge lasers have also been an area of growing interest that show promise for the realisation of optoelectronic integration. Although not pure Si lasers like the other examples here, Ge is already used as a part of Si microelectronics, making this approach inherently CMOS compatible. Ge, although also an indirect gap material has a \( \Gamma \)-minimum only 136 meV above the L-minimum leading to several groups attempting to either enhance the population of the \( \Gamma \)-minimum or form a direct gap material.

The approach of J. Liu [75] has been to use a combination of tensile strain to reduce the \( \Gamma \)-L minimum offset and n-type doping to increase the L-minimum band filling to cause a fraction of injected carriers to fall into the \( \Gamma \)-minimum, increasing the radiative efficiency. A device lasing at room temperature with 0.29 % tensile strain and \( 10^{19} \ cm^{-3} \) phosphorus doping has been demonstrated by this approach suggesting it may have potential. An alternative method for developing a Si-Ge based laser published recently is by the addition of Sn to form a direct gap material [138]. Here a direct gap, partially relaxed GeSn/Si laser is demonstrated to lase up to 90 K for a Sn concentration of 30%.

These approaches however, share a series of common issues. They are optically (other than the GeSn lasers), as opposed to electrically pumped and so require an additional light source, making their inclusion in an OEIC somewhat redundant. In addition, strong free carrier absorption reduces gain [57], saturation of the luminescence at high power operation largely due to Auger recombination [114] and for LD-Si a strong dependence on feature size of emission energy, causing significant inhomogeneous broadening and optical losses [31]. The result of these factors causes gain to be orders of magnitude below competing III-V based lasers and electrical injection lasing to be, as yet, far from realisation.

1.2.2 III-V integration on silicon

The advantage of integrating III-V active materials on Si is that these have been studied and optimised as gain media for decades and efficient lasers developed for a range of materials. The III-V can be used exclusively for its advantageous light generation properties and the Si for its excellent waveguiding and electrical data processing properties. Two main sub-categories of III-V integration on Si have been developed in order to
overcome the lattice and thermal expansion coefficient mismatch and polar/non-polar junction. Monolithic integration of materials by direct epitaxy, which can accommodate the large strain and defect densities induced by these differences. And hybrid integration, by which the III-V and Si wafers are fabricated separately and then fused by various means later, with any defects generated as a result confined to the bonding interface.

1.2.2.1 Hybrid integration  Several advantageous characteristics of the hybrid integration approach over direct epitaxy of III-Vs on Si have led to swift progress in this field. Since mature material systems are integrated onto Si that have been grown according to standard processes, comparatively little additional work is required in terms of structural or compositional optimisation of the III-V section. Of the individual approaches, the main difference affecting the III-V region is where precisely the optical mode is confined and the manner in which it is guided.

If confined to the III-V region, micro-disks allow a large optical confinement factor to be achieved, which in turn leads to increased modal gain. Micro-disks are generally formed post-bonding by reactive ion etching (RIE) of the III-V material and benefit from a small form factor that reduces overall package dimensions. The emission wavelength is determined by the diameter of the structure, allowing several of varying radius to be defined in clusters for use in wavelength division multiplexing, the drawback however is that if the radii are not repeatably achieved with suitable precision the WDM system will fail to operate correctly. Lasing from a 7.5 $\mu$m diameter micro-disk laser has recently been demonstrated with a threshold current of only 0.33 mA, however with injection raised by a factor in excess of 10 to 3.8 mA only $\sim$20 $\mu$W of optical power was coupled into the waveguide [126].

The alternative is to design the bonded device in such a way as to strongly couple the optical mode into the SOI waveguide, where specifically designed wavelength selective features are included to form a cavity resonator centered on the III-V active region spatially. This could take the form of a distributed feed back (DFB) structure for example, defined using CMOS compatible processes such as electron beam milling. In this case the optical mode couples evanescently to a III-V mesa bonded in close proximity to the waveguide where it is amplified, yielding substantially lower optical confinement factor and gain. Using this approach a larger structure is required in order to maximise the volume of active material with which the optical mode overlaps (gain volume) and as a result threshold current densities are larger. Threshold currents of $\sim$20 - 30 mA have been demonstrated with associated optical output powers in the region of 1 mW [137], giving an efficiency in excess of 6 times that of the micro-disk discussed above. The choice of cavity if therefore a tradeoff between form factor, threshold current, optical power and ease of tuning.

Several methods of hybrid integration have been demonstrated including transfer printing [54], anodic bonding [98], self-aligned solder bumps [48], adhesive bonding [165] and low temperature oxygen plasma assisted flip-chip wafer bonding [63]. At their core these approaches all have a basis in those already adopted by commercial CMOS processing and are thus generally reliable and suitable for OEIC fabrication.

One limiting factor of these methods however, is a reduced conduction of heat from the III-V active region
away to the substrate which can cause internal temperatures to rise and both efficiency and lifetime to fall. Transfer printing, anodic bonding, self-aligned solder bumps and adhesively bonded lasers are formed by first dicing the III-V wafer into individual devices and bonding either singularly or in batches to the prepared SOI wafer. The adhesive method has the advantage over flip-chip wafer bonding of far looser tolerances, and provided a suitable adhesive is chosen (such as benzocyclobutene (BCB)), strong, low optical loss and thermally stable bonds can be formed. The dies must be aligned to a precision of within a micron, be that passively (through mechanical stops for example) or actively (by manual alignment with marks or similar) and bonded. This process is inherently time consuming and pushes up the cost per unit since micro-electronics prices are generally driven by supply and time at a fabrication plant is expensive. One advantage of such a method though is a minimal consumption of III-V material, which is far more costly than Si.

Oxygen plasma assisted flip-chip wafer bonding on the other hand fuses the entire III-V wafer directly to the SOI wafer, removing the need for time consuming alignment steps, with III-V processing taking place post bonding in situ. The flip-chip wafer bonding method essentially uses Van der Waals forces, forming a strong direct bond between the III-V and Si but requires near atomically smooth surfaces (∼0.5 nm tolerance) that are totally free of contamination. Further details of the oxygen plasma assisted flip-chip wafer bonding methodology as well as a detailed study of AlGaInAs/InP/Si QW lasers formed using this process are presented in chapter 4.

1.2.2.2 Epitaxy of III-V active regions on silicon

The epitaxial growth of III-Vs on Si is the theoretically ideal solution, as fabrication would be both simplified and expedited, and a mechanically stable, defect free and reliable device could be integrated directly onto the SOI substrate. The issues discussed in section 1.1 as well as the stringent requirements of CMOS compatibility however make the realisation of such monolithic solutions challenging. The materials adopted by the telecommunications industry based on GaAs and InP, which may be readily exploited through hybrid integration cannot be directly deposited on Si without considerable crystal defects and even macroscopic cracking developing.

It has been shown [90] that III-V-Sb materials exhibit some promise in managing defect propagation from the III-V/Si interface. It is found that strain is mostly compensated through the generation of Lomer-Cottrell dislocations [60], which are immobile in the slip plane and so tend to propagate in the plane of the III-V/Si interface, generating an interfacial misfit array [128]. The result is an effective blocking of the propagation of further threading dislocations in the slip plane, allowing high crystalline quality material to be grown in subsequent layers.

The groups of Huffaker and Tourme have developed GaSb and GaInSb MQW lasers on Si demonstrating electrical injection lasing up to 77 K and 318 K with threshold current densities of 2 kAcm⁻² and 2.8 kAcm⁻², respectively [4], [89]. One issue that may need addressing in both cases is the requirement of a 5° and 4° miscut from the (100) direction of the Si surface, outside the usual maximum 0.5° required for CMOS compatibility. Nevertheless achieving lasing at 1550 nm through monolithic integration on Si is an impressive achievement suggesting there is promise for the realisation of an efficient electrical injection laser
on Si through the use of III-V-Sbs. This remains the subject of ongoing research. Details of the approach undertaken by Tournie et al. at Université Montpellier as well as an investigation of the efficiency limiting processes of such devices is presented in chapter 7.

In order to use a proven and well optimised material system the lattice constant mismatch and resultant strain must be accommodated. However, the active region and injection path can be isolated from the defects through precise strain management, which causes the bending and subsequent annihilation of threading dislocations. Several approaches have been demonstrated, including III-Sb buffer layers [4], strained Stran-skii-Krastanow quantum dot layers [104], graded Si-Ge buffer layers [97] and strained super-lattices [139]. The work of Liu et al. at UCL, for example, has recently demonstrated electrical injection 1.3 μm InAs/GaAs quantum dot lasers on Si operating up to 111 °C with a threshold current density of 200 Acm⁻² and output power of 100 mW at room temperature [144]. In addition InAs quantum dots embedded in a GaAs/AlGaAs separate confinement heterostructure on Si have been developed by the group of J. Bowers and grown by MBE. These were found to lase under continuous wave electrical injection at room temperature with a threshold current density of ~0.5 kAcm⁻². Thermal rollover limited the maximum operating temperature to 120 °C. Single facet wall plug efficiencies of 18%, operation at 1.3 μm and output power of ~176 mW at ambient temperature indicate that these devices have potential for realisation of a viable silicon compatible laser [10].

The work of Gossard’s research group [150] demonstrated that quantum dots may also be used as an active region in order to achieve lasing on Si. If it is assumed that there will be a certain defect density when growing III-Vs on Si, rather than looking to avoid defect generation the approach could be to simply prevent carriers reaching the defects. If a strong enough quantum confinement potential is achieved, in this case through the use of InAs/GaAs quantum dots, then it may be possible to avoid carriers recombining with defects and therefore eliminating their negative impact on device performance [150]. Threading dislocations can be problematic in quantum well structures owing to the fact that carriers with high in-plane mobility will concentrate at the resulting defect and recombine non-radiatively. In quantum dot systems, this is less problematic owing to the fact that the carriers occupy spatially separated gain regions meaning that while a dislocation may inhibit gain in a given quantum dot, it need not affect the gain in other quantum dots.

An alternative approach investigated extensively at Philipps University, Marburg is the direct epitaxial growth of the dilute nitride GaNAsP on silicon using the scalable MOVPE growth approach. Since GaP exhibits a very similar lattice constant to Si, defect free growth of GaP on Si has been readily achievable and demonstrable for some time under the correct growth conditions [49]. GaP however is an indirect gap material like Si and so is also unsuited to efficient photon generation. The addition of large As fractions forms a direct gap ternary, however the lattice constant mismatch is increased considerably. In order to form a direct gap material lattice matched to Si nitrogen in added to form the dilute nitride GaNAsP, which meets the aforementioned criteria for ~ 4% N and > 70% As fractions [20]. In addition to this active material, novel strain compensating boron containing materials have been developed to manage the strain induced during growth by the differing thermal expansion coefficients. The materials $B_{0.033}Ga_{0.967}P$ and $(B_{0.05}As$ are found to act as both suitable strain compensation layers as well as SCH and barriers respectively. Studies of the
fundamental properties of these novel materials and their optimal growth conditions has been required in order to achieve defect free growth [109],[112]. Electrical injection lasing of a GaNAsP/BGaP single quantum well laser monolithically grown on a silicon substrate has been demonstrated up to a maximum temperature of 165 K has been achieved with a threshold current density of $1.6 \, kAcm^{-2}$ and emission wavelength of 861 nm [108]. Greater detail on the development of the GaNAsP material system, its integration on silicon, past studies of efficiency limiting processes and studies of these processes for different active region structures and Si surface orientations can be found in Chapter 5. Chapter 6 presents a theoretical study of the optical and electrical properties of these devices as a part of developing a full active device simulation and for optimisation of optical confinement. Finally this chapter also includes a study of processing methodology optimisations including lithography, etching and metalisation investigations as well as a characterisation of the devices produced by the optimised method in comparison to previous devices.

1.3 Thesis aims

Sections 1.2 indicates that the III/V integration on silicon as oppose to lasers developed from silicon or Ge have shown the most promise thus far. Therefore it is this approach that is the focus of this thesis. Of the four most developed approaches; monolithic epitaxial growth of the dilute nitride GaNAsP/BGaP, hybrid integration, direct epitaxial growth of III/Sbs and quantum dot based lasers samples of all but the QD based lasers were obtained for investigation. The facilities and expertise at the University of Surrey make experimental characterisation of devices as a function of temperature and particularly high hydrostatic pressure. Thus, there is a clear opportunity for this thesis to uniquely bring together directly comparable studies of the lasing characteristics and physical processes, which limit the efficiency of all but one of the leading III/V based approaches to optoelectronic integration in a single document. With this in mind the aims of this thesis are as follows:

- Identify, and where possible quantise the contribution of non-radiative recombination mechanisms, which limit efficiency, increase threshold current density and limit operating temperature and output power. This is primarily achieved through temperature and pressure dependent electroluminescence, observing both the emission from the facet and pure spontaneous emission collected from the substrate.

- Identify and where possible quantise any other physical processes and properties, specific to each method of integration on Si that may limit performance such as optical losses, optical confinement factor, diffusion of impurities, processing quality etc. Methods of investigation will be specific to observed or hypothesised mechanisms and properties, and will be chosen as required to suit (for example simulation using Rsoft in the case of optical confinement factor or SEM for metalisation and etch quality).

- Suggestion of possibilities for improvements and optimisations of devices based on the results of the previous two aims. The details of growth techniques are not a focus of this work and so if for example band engineering based solutions were suggested, then nominal changes and potentially material compositions that may be suitable would be the expected level of details provided.
• Characterisation of devices based on suggested changes would allow; firstly the results and conclusions on which the changes were based to be tested, and secondly the impact of those changes on performance and limiting processes to be studied. Thus, by such an iterative process the devices should gradually tend towards a commercially viable laser as limiting processes are minimised and beneficial properties such as modal gain and optical confinement factor maximised.

• Overall to produce a single document which summarises the state of the field of III/V integration on Si, highlighting the benefits and shortcomings of the competing approaches.

1.4 Thesis outline

• Chapter 1: The background, motivation and requirements for pursuing optical integration on Si is presented along with a summary of the current state of the field for the main approaches to the realisation of a silicon compatible laser.

• Chapter 2: The basic theory relevant to the work presented in this thesis is outlined including semiconductor band structure, the requirements for achieving lasing, efficiency limiting processes, characterisation of semiconductor lasers and the effects of temperature and pressure on these devices. Additional more specific theory is presented at the opening of the relevant chapters.

• Chapter 3: The experimental methods used to collect the data presented chapters 4 - 7 are described including the equipment, methodologies and techniques required to carry out each type of study.

• Chapter 4: An experimental investigation into the efficiency limiting processes of the hybrid AlGaInAs silicon-evanescent lasers produced by the group of J. Bowers at UCSB is presented. The dominant sources of loss are determined as a function of temperature for distributed feedback (DFB) waveguide lasers of two types; with and without a strained superlattice defect-blocking layer, and the performance of each compared. Devices with a Fabry-Perot cavity and optimised bonding procedure are also characterised as a function of temperature and the performance of these compared to the DFB waveguide devices.

• Chapter 5: The effect on device performance and efficiency limiting processes of having multiple as oppose to single quantum wells is investigated as a function of both temperature and pressure for GaNAsP based lasers monolithically integrated on silicon. A similar study is also carried out for devices which have a differing Si substrate orientation to determine the effect, if any on performance in order to further optimised the growth of the GaNAsP based material system. By combining the results of both temperature and pressure dependent electroluminescence studies, the fractions of injected current associated with each recombination mechanism are quantified and a fundamental understanding of the limitations of the devices gained.

• Chapter 6: A passive waveguide simulation of the GaNAsP based lasers is carried out using the Laser-Mod application from Rsoft. By optimising the thicknesses of the barrier and cladding layers the optical
confinement factor and modal gain are increased without sacrificing other parameters such as carrier confinement. Secondly, a study of the processing methodology for the GaNAsP/BGaP material system is detailed. Including the optimisation of lithography, etching and metalisation in order to improve on existing processing methods.

- Chapter 7: An experimental study of the efficiency limiting processes of GaSb based lasers for monolithic integration on silicon as a function of both temperature and hydrostatic pressure is presented. The study identifies and quantifies the fractions of each efficiency limiting process taking place in these devices for the purposes of optimising the lasers before integration on silicon.

- Chapter 8: The conclusions of each chapter are brought together in a summary and suggested further studies based on these findings are presented.
Chapter 2

2 Theory

Chapter 2 summarises the theory required to understand the subsequent material within this thesis. It relates to the general background of semiconductor lasers with a focus on recombination mechanisms and efficiency limiting processes. More specific theory is also presented at the opening of each results chapter as required. It is assumed that the reader has some understanding of basic solid state physics including crystal structure in addition to quantum mechanical concepts such as the Pauli-exclusion principle, quantum confinement and wave-particle duality.

2.1 Semiconductor fundamentals

2.1.1 Electronic band structure

The periodic nature of a perfect semiconductor crystal translates corresponding periodicity and discretisation to the electron wavefunctions in the form of Bloch functions [88]. Wavefunctions of this type lead to the formation of bands of allowed energies separated by so called band gaps ($E_g$). The periodicity of the band structure continues indefinitely but only a finite number of electrons are present within the crystal, which to minimise the energy of the system fill the lowest energy states first, with only one electron spin per state according to the Pauli exclusion principle [118].

Thus at some energy the filling will cease and the subsequent bands will be empty, in the case of a semiconductor at 0 K the uppermost valence band is totally full with the next band completely empty. However, at greater temperatures the thermal energy can be great enough to allow some electrons to gain energy in excess of the band gap and pass into the conduction band, leaving both partially filled. The probability of states being filled as a function of temperature is defined by the Fermi-Dirac distribution, as detailed in section 2.1.3. The band into which the thermally excited carriers are excited is known as the conduction band (CB), whilst the band from which they came is called the valence band (VB). The valence band is in turn split into three sub bands known as the heavy hole (HH), light hole (LH) and spin-orbit (SO) bands. An additional forbidden energy gap exists between the LH and SO bands known as the spin-orbit splitting and can become relevant when discussing non-radiative recombination mechanisms or optical losses. This arrangement of bands is depicted for a simplified case in figure 2.1 and is used as a basis for the discussion of electron states and transitions.
The energies of these bands are defined as a function of wave vector according to a parabolic band approximation [37]. This holds close to $k = 0$, known as the Γ-point, which is the region of interest for the bulk of situations considered during the discussion of semiconductor lasers giving an energy of the form

$$E = \frac{\hbar^2 k^2}{2m^*}. \quad (2.1)$$

Equation 2.1 indicates that the band energies are dependent upon the wave vector, $k$, Planck’s reduced constant, $\hbar$ and $m^*$, a parameter known as the effective mass. The curvature of the dispersion relation defines the effective mass of the carriers which occupy the band, in the case of the conduction band the curvature is positive and so is $m^*$, however for the valence bands it is negative, yielding a negative $m^*$. The existence of a negative electron mass is somewhat unphysical, leading to the definition of the “hole” as a new particle with the same properties as an electron but with negative mass.

The physical reality is that no such particle exists, but can be rationalised if one considers the movement of $(n-1)$ electrons in one direction through $n$ states under an electric field. It is totally equivalent, but far simpler to describe the motion of the one vacant state in the opposite direction, this vacancy is what the hole particle represents. Therefore this simple model defines a system of bands in which the three valence are mostly full of electrons with “holes” towards to the top. Whilst the conduction band is mostly empty with electrons towards the bottom. The minority particle in each case is known as a charge carrier, since it is able to move.

The band gap, determined by the periodicity and lattice spacing of the crystal is sensitive to temperature changes, since these will cause the crystal to expand and contract, varying the lattice constant. The temperature dependence of the bandgap is found to be well predicted using the Varshni coefficients, $\alpha$ and $\beta$ and the band gap at 0 K, $E_g(0)$ according to the following empirical equation:
\[ E_g(T) = E_g(0) - \frac{\alpha T^2}{T + \beta}. \] (2.2)

The parameters to be used in equation 2.2 are determined experimentally and can looked up from literature for most common semiconductor materials and their binaries. The parameters for more complex ternary and higher order alloys can be determined by interpolation of the values for the constituent binaries weighted by the concentrations of each required to form the final material.

### 2.1.2 Bulk density of states

Each of these bands contains multiple states and as can be seen from figure 2.1 the number increases with both effective mass and energy. Thus to describe these states we define the parameter known as the density of states [88], \( \varphi \), as given by equation 2.2 for a bulk semiconductor material:

\[ \rho = 4\pi \left( \frac{2m^*}{\hbar^2} \right)^{3/2} \times E^{(1/2)}. \] (2.3)

The term \( E \) is defined by equation 2.1, and defines the energy measured from the band edge into the band at which the density of states is to be calculated, with the band in question determined by the effective mass. Through equation 2.1 the effective mass is related to the wave vector, which in turn is dependent on the lattice constant (separation of ions within the lattice). The effective mass and density of states therefore vary for from material to material, strongly influencing both the optical and electrical properties. It can be seen from equation 2.3 that the density of states scales according to the square root of the energy from an initial value of zero at the band edge.

### 2.1.3 The Fermi-Dirac distribution

As was mentioned above the Pauli exclusion principle applies to carriers in semiconductor materials since they are fermions and so once each state is occupied the next highest in energy must be filled subsequently, however additional thermal energy will cause a distribution of filled states around the valence band maximum and conduction band minimum. Thus a definition of the probability of a state being occupied is defined, known as the Fermi-Dirac distribution [88] and is given by equation 2.4:

\[ f(E) = \frac{1}{1 + \exp \left( \frac{E - E_f}{k_BT} \right)}, \] (2.4)

where \( f(E) \) is the probability a state of energy \( E \) would be occupied, \( k_b \) is Boltzman’s constant, \( T \) is the temperature in Kelvin and \( E_f \) is the Fermi-energy, the value of energy that defines the point at which the probability of occupation is 50%. The distribution is plotted in figure 2.2 for 0, 150 and 200 K with an arbitrary Fermi-energy of 200 meV. The occupation probability at \( 4k_BT \) (~2%) is also marked as this is a common approximate upper limit for carrier thermal energy.
Figure 2.2: The Fermi-Dirac distribution for a Fermi-energy of 200 meV and temperatures of 0, 150 and 300 K. The occupation probability at $4k_B T$ above $E_f$ falls to approximately 2% and is often used to mark the approximate upper limit of carrier energy.

At thermal equilibrium $E_f$ is located at the centre of the band gap, since all electrons in the conduction band transitioned from the valence band, which would otherwise be full. The probability for hole occupation in the valence band is then just $(1 - f(E))$, since a state either has an electron or a hole in it. However, in non-thermal equilibrium situations where electrons or holes are added to their respective bands from some external source such as doping the Fermi-level may move, or even split into two separate quasi-Fermi levels under electrical current injection. The quasi-Fermi levels $E_{fc}$ (conduction band Fermi-level) and $E_{fv}$ (valence band Fermi-level) represent those situations for which the system is not in thermal equilibrium and a single Fermi-level therefore cannot describe both electron and hole occupation probabilities. Since there are two Fermi-levels, one for each band, each distribution is calculated separately giving $f_c$, the occupation probability for conduction band states and $f_v$, the occupation probability for valence band states.

2.1.4 Carrier densities

With both the density of states and probability of those states being occupied known the product of the two will give the density of carriers at each given energy, integrating from the band edge to infinity then gives the density of carriers over all energies. The electron and hole densities in the conduction and valence bands respectively are given by equations 2.5 and 2.6 below.

\[
n = \int_{E_c}^{\infty} \rho_c(E) f_c(E) dE \tag{2.5}
\]
\[ p = \int_{-\infty}^{E_v} \rho_v(E) (1 - f_v(E)) \, dE \quad (2.6) \]

Here \( n \) and \( p \) are the electron and hole densities, \( E_c \) and \( E_v \) are the energies of the conduction and valence band edge energies, \( \rho_c(E) \) and \( \rho_v(E) \) are the conduction and valence band densities of states at energy \( E \) and \( f_c(E) \) is the electron occupation probability at energy \( E \) in the conduction band (making \( 1 - f_v \) the hole occupation probability in the valence band). With these equations it is now possible to define the density and distribution of charge carriers throughout the band structure of semiconductor materials as a function of both energy and temperature. These will be of use in describing carrier dynamics and the recombination of carriers, the key processes that define the performance of a semiconductor laser.

### 2.2 Radiative electronic transitions

Electrons may transition between the states discussed in section 2.1, but if the states differ in energy the excess or deficit must be accounted for by interaction with another particle. Intraband transitions tend to take place via phonon emission in order to conserve momentum \((h\mathbf{k})\), since photons carry little momentum and states of differing energy within a band have different \( k \) values. These transitions represent processes like scattering and are classed as non-radiative since they do not involve photons. Interband transitions may however take place between states of identical wave vector, removing the constraint on momentum and therefore may take place via the exchange of photons, classing them as radiative transitions. The three radiative transitions are depicted in figure 2.3.

![Figure 2.3: The three radiative transitions that may take place in a direct band gap semiconductor. Left: Photon absorption. Centre: Spontaneous emission. Right: Stimulated emission.](image)

#### 2.2.1 Absorption

Absorption (figure 2.3 left) describes the process of a valence electron with energy \( E \) interacting with an incident photon with sufficient energy \((h\nu)\) to allow a transition beyond the band gap and into a vacant conduction band state (at energy \( E + h\nu \)) to take place, leaving a hole in the valence band. This process generates two free carriers that may now move within their bands since there are an abundance of empty states
for them to translate into, allowing a current to flow. The rate of absorption depends on three parameters, the density of electrons of energy E within the valence band \((\rho_v(E)f_v(E))\), the density of empty states in the conduction band at energy \(E + h\nu \ (\rho_c(E + h\nu)[1 - f_c(E + h\nu)])\) and the density of photons with energy \(h\nu \ (P(E_{h\nu}))\) (where the condition \(h\nu \geq E_g\) must be met). Multiplying these parameters together along with a constant of proportionality known as the Einstein \(B_{12}\) coefficient gives equation 2.7,

\[
  r_{abs} = B_{12}\rho_v(E)f_v(E)\rho_c(E + h\nu)[1 - f_c(E + h\nu)]P(E_{h\nu}).
\]  

(2.7)

2.2.2 Spontaneous emission

Spontaneous emission (figure 2.3 centre) is the opposite process to absorption, where an electron in the conduction band with energy \(E + h\nu\) transitions to a vacant state in the valence band (hole) of energy \(E\) and in doing so emits a photon of energy \(h\nu\) to conserve energy. The name indicates that the process is spontaneous, which is to say a conduction band electron interacts with a virtual photon causing a de-excitation into an empty valence band site. The photon emitted will therefore have any allowed energy, \(h\nu\) for which a transition is possible and will travel in a random direction. The rate is dependent upon the density of electrons with energy \(E + h\nu \ (\rho_c(E + h\nu)f_c(E + h\nu))\) and the density of holes with energy \(E\) in the valence band \((\rho_v(E)[1 - f_v(E)])\) only and multiply together along with the Einstein \(A_{21}\) coefficient to give the spontaneous emission rate:

\[
  r_{sp} = A_{21}\rho_c(E + h\nu)f_c(E + h\nu)\rho_v(E)[1 - f_v(E)].
\]  

(2.8)

The radiative current, \(I_{rad}\) (the fraction of the injected current that is associated with radiative recombination) is found to be proportional to the the spontaneous emission rate [148]; This assumes that \(n = p\), which is a reasonable assumption under electrical injection at an LED or laser heterojunction/quantum well,

\[
  r_{sp} \propto I_{rad} = eVBn^2.
\]  

(2.9)

Where \(e\) is the elementary electronic charge, \(V\) is the volume of the electrically pumped active region, \(B\) is the radiative coefficient (and related to \(A_{21}\)) and \(n\) is the carrier density (assumed equal for conduction electrons and valence holes).

2.2.3 Stimulated emission

Stimulated emission (figure 2.3 right) is similar to spontaneous emission, but recombination is “stimulated” by the interaction of a photon with energy \(h\nu\) and a CB electron. When this takes place a photon of energy \(h\nu\) is still emitted to conserve energy, but this time is emitted with an identical direction and phase to the original photon, giving two photons that are completely in phase and overlap spatially. This process is what allows a laser to operate, since it causes the photon density at a “single” energy to be amplified and is required to achieve net gain (amplification) as photons pass through the material. The similarity to spontaneous emission
is reflected in equation 2.10, which is essentially 2.8, but with the additional dependence on the density of photons with energy $h\nu \left( P(E_{h\nu}) \right)$ and the Einstein $A_{21}$ coefficient replaced by the Einstein $B_{21}$ coefficient.

\[
r_{st} = B_{21}\rho_c(E + h\nu)f_c(E + h\nu)\rho_v(E)|1 - f_v(E)|P(E_{h\nu}).
\] (2.10)

### 2.3 Non-radiative carrier recombination

Stimulated emission is the only desired carrier recombination for a laser and so it is this that must be maximised. However, in addition to desirable stimulated emission there are several non-radiative recombination mechanisms (as well as spontaneous emission) that will consume current. Since these processes deplete free carriers but generate no photons (generally phonons/heat instead) they limit the efficiency of the laser and are therefore undesirable and should be minimised wherever possible.

#### 2.3.1 Defect-related recombination

Thus far a prefect crystalline lattice has been considered, however the reality of crystal growth is that some imperfections will always be present. These imperfections can take the form of material impurities, vacant lattice sites, dislocations of atoms from lattice sites and interstitial inclusions. Regardless of their type or origin, the basic action of a defect is to interrupt the perfect periodicity of the lattice and therefore cause localised disruptions to the band structure. Since the defects do not conform to the periodicity of the lattice the host band gap does not apply to their states and they may form within the bandgap (see figure 2.4).

Carriers may become trapped in these states, or, if there are several states of differing energy then carriers may undergo several non-radiative transitions in order to recombine. The result is a loss of injected carriers and the generation of excess phonons/heat, which can cause other non-radiative recombination mechanisms to become dominant. Defects are also not necessarily static and may irreversibly move through the structure as a result of strain, elevated temperature or an applied electrical bias. Thus by this mechanism the quality of a laser may degrade over time, limiting efficiency and operational lifetime.
Recombination via defects, also known as Schockley-Read-Hall (SRH) recombination or monomolecular recombination is dependent only on the defect density and carrier density, meaning the defect current is given by:

\[ I_{SRH} = eVAn, \]  

where \( A \) is the monomolecular recombination coefficient (not to be confused with \( A_{21} \), the Einstein A coefficient) and is a function of the defect density, carrier capture cross section and thermal velocity of carriers.

The SRH recombination mechanism is only weakly dependent on temperature (compared to those it competes with) and is minimised mainly through the optimisation of crystal growth, in order to achieve high quality crystalline material. Therefore in mature material systems with well-developed growth procedures defect-related recombination can be assumed to be negligible.

### 2.3.2 Auger recombination

Auger recombination is a three carrier Coulombic interaction leading to an exchange of energy and momentum. The net effect is an interband recombination accompanied by an intraband carrier excitation with energy and momentum being conserved. There are several types of Auger recombination that vary according to the arrangement of the carriers. The main two of relevance to this study are conduction to heavy hole exciting a conduction electron to a higher energy conduction band state, or CHCC for short and conduction to heavy hole exciting a spin-orbit electron to the heavy hole band or CHSH, both of which are depicted in figure 2.4. The other Auger processes, CHLH and CHHH are less impactful in the lasers discussed in this thesis since the CHLH process has been shown to be \(~100\) times weaker than either CHCC or CHSH, with CHHH weaker still due to the high effective mass of the carriers involved [146].
Auger recombination is therefore a three-body process dependent on the conduction electron density, heavy hole density, and third carrier density, which will vary depending on the type of Auger process being discussed. Thus all Auger processes are approximately proportional to the cube of the carrier density (again assuming $p = n$), giving an Auger current of [147]:

$$I_{Aug} = eVCn^3,$$  \hspace{1cm} (2.12)

where $C$ is the Auger coefficient, which is band-structure, temperature and pressure dependent. Auger recombination is therefore a recombination mechanism that can be both exacerbated and suppressed by a range of factors. If CHCC and CHSH are assumed to be the two most relevant forms of Auger recombination and again that $n = p$, then the the Auger coefficient is defined as follows:

$$C = C_0^{CHSH} \exp\left(-\frac{E_a^{CHSH}}{kT}\right) + C_0^{CHCC} \exp\left(-\frac{E_a^{CHCC}}{kT}\right).$$  \hspace{1cm} (2.13)

Splitting $C$ into two separate factors gives two temperature independent constants, $C_0^{CHSH}$ and $C_0^{CHCC}$ and defines the terms $E_a^{CHSH}$ and $E_a^{CHCC}$, known as the Auger activation energies for the respective mechanisms. The exponential term indicates that the Auger coefficient varies strongly with temperature. When considering equation 2.12, the temperature sensitivity of the Auger coefficient in addition to the cubic dependence on the carrier density causes the Auger current to be highly dependent on the temperature. The activation energies for the CHSH and CHCC Auger process are given by equations 2.14,

$$E_a^{CHSH} = \frac{m_{SO}}{2m_h + m_e - m_{SO}} (E_g - \Delta_{SO})$$  \hspace{1cm} (2.14)

and 2.15,

$$E_a^{CHCC} = \frac{m_e}{m_h + m_e} E_g$$  \hspace{1cm} (2.15)
and are dependent on both the effective masses of the bands involved in the process as well as band gap or band gap to spin-orbit splitting energy separation in the cases of CHCC and CHSH respectively. The temperature dependence of the band gap then adds another source of temperature sensitivity that affects both types of Auger mechanism. In the case of CHSH Auger recombination there is an additional dependence on the difference between the band gap and spin-orbit splitting that will scale the activation energy of the process as can be seen in figure 2.6.

![Figure 2.6: The trend in normalised CHSH Auger recombination coefficient with bandgap - spin orbit splitting difference and temperature.](image)

In the limit $E_g \gg \Delta_{SO}$ the activation energy is large and $C$ is small leading to minimal Auger recombination, $E_a^{CHSH}$ increases linearly as $E_g \rightarrow \Delta_{SO}$ with $C$ becoming significant as $E_a^{CHSH} \rightarrow k_bT$. If the condition of $E_g = \Delta_{SO}$ is met then $E_a^{CHSH} = 0$ and the CHSH Auger process becomes resonant. However, if the band gap is less than the spin orbit splitting then CHSH can be suppressed, making the processes effectively impossible in most practical situations.

This is clearly beneficial for near and particularly mid-infrared and longer wavelength emitters, which are known to be limited by CHSH Auger recombination due to their small band gap [81]. Increasing the spin-orbit splitting is a case of band engineering by adjusting the material composition during growth. The addition of a small fraction of bismuth has shown promise in recent studies for achieving CHSH Auger suppression in GaAs and InP based material systems for use in the near infra-red [69]. The large $\Delta_{SO}$ of some antimonide alloys has also demonstrated suppression of CHSH Auger recombination in mid-infrared lasers [132]. The temperature dependence in the exponential reflects the greater thermal distribution of carriers in both energy, and particularly wave vector since momentum must be conserved and so a greater range of wave vectors allows
for more possible “hot” carrier generation possibilities.

Auger recombination then is a highly temperature dependent mechanism as $I_{\text{Aug}} \propto n^3$ and $n \propto T$, the Auger coefficient, $\propto \exp \left( \frac{-E_{\text{a}}}{T} \right)$ and the activation energy has additional temperature dependence inherited from the band gap. This mechanism can be identified through temperature dependent studies and can be assumed to be negligible at very low temperatures (less than \(~\sim 100\) K), whilst likely being one of the limiting factors in operating temperature when this process is present to any reasonable degree.

### 2.3.3 Carrier leakage

Carrier leakage is the process of carriers injected into the device junction/quantum well passing through without undergoing recombination in the active region. In order to keep carriers within the locality of the junction or quantum well, barrier materials with a larger band gap are used to create a potential barrier that confines the carriers so that they stay in the active region for longer. Figure 2.7 depicts a quantum well structure from which a proportion of carriers has sufficient energy to escape, overcoming the energy barrier.

![Figure 2.7: A schematic diagram of carrier leakage from a quantum well.](image)

The leakage current is dependent on the density of carriers with energy sufficient to overcome the potential barrier, their mobility and lifetime, as is shown in equation 2.16:

$$I_{\text{leak}} = \frac{eVDn_{\text{leak}}}{L} \quad (2.16)$$

$D$ is the diffusion coefficient (proportional to the mobility), $L$ is the diffusion length (proportional to the square root of the carrier lifetime) and $n_{\text{leak}}$ is the carrier density with energy exceeding the potential barrier of the active region. $n_{\text{leak}}$ is defined using the same process as the total carrier density, integrating the product of the density of states and Fermi-Dirac distribution for the conduction band, but this time only integrating from above the height of the confining potential, $E_B$ to infinity.
\[ n_{\text{leak}} = \int_{E_B}^{\infty} \rho_c f_c dE. \]  
(2.17)

Assuming \( E_B \gg k_b T \) Boltzmann statistics are a suitable approximation, hence we can state:

\[ n_{\text{leak}} \propto n_0 \exp \left( -\frac{E_B}{k_b T} \right). \]  
(2.18)

and therefore:

\[ I_{\text{leak}} = I_0 \exp \left( -\frac{E_B}{k_b T} \right). \]  
(2.19)

It should be noted that this assumes a temperature independent diffusion coefficient for simplicity, giving a temperature dependence as follows:

\[ I_{\text{leak}}(T) \propto \exp \left( -\frac{E_B}{k_b T} \right). \]  
(2.20)

Carrier leakage has a large temperature dependence and will often limit the thermal characteristics of LEDs and lasers. Once carriers escape the active region they may undergo either radiative or non-radiative recombination processes, but in either case will not contribute to “useful” photon production. Hole leakage is generally significantly less than electron leakage due to the higher effective mass and therefore lower mobility as well as the fact that the quasi Fermi-energy for holes is usually above the valence band maximum.

The total injected current can therefore be defined in terms of components of each recombination mechanism as follows:

\[ I = eV (An + Bn^2 + Cn^3) + I_{\text{leak}}. \]  
(2.21)

Where \( e \) is the electronic charge, \( V \) is the pumped volume, \( n \) is the carrier density and \( A, B \) and \( C \) are the recombination coefficients for defect, radiative and Auger recombination, respectively. An additional term, \( I_{\text{leak}} \) is added to take account of carrier leakage from the active region. This relation shows that the higher the threshold carrier density the greater the contribution due to Auger recombination and (due to greater band filling) carrier leakage.

### 2.4 The pn junction

The basic semiconductor structure that the laser is based on is the pn junction, where \( p \) and \( n \) refer to semiconductor materials that have been doped with atoms with a lower or higher valence than the host respectively. The dopants are effectively crystal defects that have been intentionally included for the purpose of increasing the minority carrier density in the valence or conduction band for \( p \) and \( n \) type materials respectively. The effect on the band structure is illustrated in figure 2.8, the p-type dopants create what is known as an acceptor defect level just above the valence band edge and the n-type dopants create a donor
level just below the conduction band edge.

The deficit/excess of one electron per atom of the p/n type dopants causes electrons/holes from the valence/conduction band edge to move into the acceptor/donor levels since the splitting from the valence/conduction band is less than $k_B T$ at ambient temperature. Thus, a p-type dopant increases the hole density in the valence band by accepting electrons into its defect level and a n-type dopant increases the electron density in the conduction band by promoting its excess electrons from its defect state into the band. Since the “intrinsic” state of all minority carriers having been generated by band-to-band transitions has been perturbed, the Fermi-level moves to match the new electron distribution, towards the valence band for a p-type material and towards the conduction band for a n-type material. It should be noted that both types of material remain electrically neutral since the dopant atoms themselves are neutral and the electrons have simply been redistributed. Typical dopant densities for use within lasers are approximately $10^{17}$ - $10^{19}$ dopants (additional carriers) per cubic centimetre.

![Figure 2.8: Band diagrams illustrating the effect of adding p (left) and n (right) type dopant atoms to a host crystal for non-zero temperatures.](image)

In thermal equilibrium there can only be one Fermi-level and so if p and n-type materials are brought into contact the bands must reconfigure at the interface to account for this, creating the depletion region. At the junction, some of the electrons from the n-type region diffuse (concentration driven) into the p-type material and vice-versa since the concentration of the opposite carrier in these regions is far lower. However, the previously electrically neutral layers now become charged due to the imbalance resulting from diffusion, causing an electric field to be formed opposing the diffusion of further carriers. At some point these two processes will balance and a steady change from n-type to intrinsic to p-type will be formed, which allows a single Fermi-level to exist throughout the structure as depicted in figure 2.9 (top).

If electrical contacts are placed on the p and n-type regions and connected to a an electrical power supply then a current should be able to easily flow in the doped region due to the high density of charge carriers, electrons flowing in from the n-type conduction band and holes in the p-type valence band. The potential barrier resulting from the doping-induced band offsets will however prevent current flow as long as the energy of the carriers is less than that of the potential barrier.

24
Therefore a turn-on voltage that allows carriers to overcome the built in potential must be achieved before a current will flow, beyond which the IV relationship will be linear. This is the characteristic IV of a diode and gives semiconductor lasers the alternative name diode lasers. Above turn on electrons injected through the n-side conduction band will meet holes injection from the p-side valence band at the centre of the depletion region causing the carrier density at this point to rise and making radiative recombination more probable. This is depicted in figure 2.9 (bottom) and causes the aforementioned Fermi-level-splitting near the junction where the system is no longer in thermal equilibrium and thus cannot be defined by a single Fermi-level. This is the basic concept on which all semiconductor LEDs and lasers are based, although many improvements can be made as will be discussed in subsequent sections.

![Unbiased and Biased PN Junction Diagram](image)

Figure 2.9: A simplified pn junction indicating a single Fermi-energy, junction potential and depletion region.

### 2.5 Gain and population inversion

#### 2.5.1 Population inversion

The purpose of a laser is to generate light through the process of stimulated emission almost exclusively, in doing so the light generated is approximately monochromatic, coherent, highly directional and of a single polarisation, as required for communications applications. Therefore, even if non-radiative recombination has been completely eliminated the laser must be designed and operated in such a way as to cause the stimulated emission rate to exceed both absorption and spontaneous emission at a given energy. Comparing equation 2.8 to 2.10 it can be seen that $r_{stim}(E)$ will increase relative to $r_{spont}(E)$ proportionally to the photon density.
with energy $h \nu$, and so if $r_{\text{stim}}(E)$ can be made to exceed $r_{\text{abs}}(E)$ then the photon density will be amplified by the stimulated emission process and it will become the dominant radiative recombination mechanism. Therefore using equations 2.7 and 2.10 the condition for optical gain is:

$$r_{\text{stim}} > r_{\text{abs}} \rightarrow f_c > f_v.$$ (2.22)

Using equation 2.4 and substituting in the quasi-Fermi levels of the conduction and valence bands gives the condition:

$$E_{fc} - E_{fv} > E_C - E_V.$$ (2.23)

$E_C$ and $E_V$ are conduction and valence band energies between which the recombination takes place, equal to the photon energy, $h \nu$ and $E_{fc} - E_{fv} = \Delta E_f$ is the quasi-Fermi level splitting. Combining this with equation 2.23 gives the Bernard-Duraourg condition for population inversion [42]:

$$E_{fc} - E_{fv} > h \nu > E_g.$$ (2.24)

Population inversion, defined by equation 2.24 represents the condition of the minority carrier density exceeding the majority carrier density at the transition energy, causing the rate of photon emission to exceed the absorption rate at that energy. This is known as population inversion since the population of carriers is reversed with respect to the thermal equilibrium state. The point at which the quasi-Fermi level splitting is equal to the photon energy is called transparency and represents the state at which, on average, a photon will pass through the full length of the laser without being absorbed. This is on the border of population inversion, where the majority and minority carrier densities are equal at $E = h \nu$.

In order to achieve the Bernard-Duraourg condition additional minority carriers must be added to both the valence and conduction bands. Doing so is known as pumping and can be achieved in two ways. Optical pumping involves injecting carries by illumination with a sufficiently intense source of photons (usually another laser) with a greater energy than the emission energy desired from the laser being pumped. This generates carriers through absorption; which thermalise until they reach the energy levels of the lasing transition where they can contribute to the population inversion. The problem with this approach is that it requires another, more powerful laser to start with, making the process as a whole less efficient and impractical for a lot of applications, including optical integration into microelectronics. The second approach, used in most commercial lasers is electrical injection, by which the laser has current injection regions (a pn junction) and electrical contacts within the design, allowing an electrical power supply to inject carriers into the device. This is more conducive to integration into other structures, gives greater overall system efficiency and minimises the footprint of the device.
2.5.2 Gain

Beyond transparency a true inversion is achieved and a pulse of light traveling the length of the laser will be increased in intensity, rather than lowered, as would be the case in the un-pumped material. This amplification of the photon density is known as optical gain and is a key defining feature of lasers. Gain spectra as a function of injection current (pumping) are plotted in figure 2.10, highlighting three separate regions of interest. For photon energies of less than the band gap the material is transparent, since the photons do not have sufficient energy to cause interband transitions. At energies between the band gap and quasi-Fermi-level splitting (those that satisfy equation 2.24) gain will be experienced to a degree dependent on the injected current density (pumping) and the carrier density at that energy (Fermi-Dirac probability and density of states product).

If the photon energy exceeds the quasi-Fermi-level splitting then the Bernard-Duraffourg condition is no longer met and absorption (negative gain) takes place. As the injection current increases the bands are filled to a greater level with minority carriers and since the density of states increases with energy so does the carrier density for a given energy, causing a blue-shift in peak gain. For Fabry-Perot cavity devices (see section 2.6.1) the lasing emission will be centered on this energy since all other photon energies will be amplified less, leading to an increased rate of photon generation at the peak energy. The emission below lasing threshold will take place over a range of energies given by the gain spectrum, as represented by figure 2.10 but is rapidly narrowed as threshold is approached, as will be explained in the subsequent section.

Figure 2.10: Gain as a function of carrier injection in a semiconductor.
2.6 The cavity and achieving lasing threshold

2.6.1 The Fabry-Perot laser cavity

Meeting the Bernard-Duraourg condition and achieving net gain, although favoring stimulated emission, does not lead to lasing in practical devices. As can be seen from figure 2.10 modal gain values of the order of $10 \text{ cm}^{-1}$ are expected, compared to $67 \text{ cm}^{-1}$ at lasing threshold for a commercial GaAs/AlGaAs laser [100]. A common length for a Fabry-Perot laser is 1 mm, meaning that the maximum amplification a photon could experience assuming it is generated at one end and emitted at the other is a factor of 812, falling exponentially to zero as the generation point moves towards the emission facet. If a conservative emission power of 1 mW is desired then $1/812 \approx 1.2 \mu W$ would have to be initially generated by spontaneous emission if we take the ideal (extremely improbable) situation of all photons being generated at the opposite end of the laser and all of them traveling along the length rather than any other direction. Assuming a photon energy of 1.4 eV (GaAs band gap) this is $\sim 5.4 \times 10^{12}$ photons per second, which taking into account the random emission direction is a total photon generation of $\sim 5.4 \times 10^{19}$ photons per second. Converting to the current required to generate this number of photons per second gives approximately 8.6 A, a somewhat impractical value for 1 mm long device.

The issue is that the devices are so short, giving the photons little time to be amplified before being lost from the facet of the laser, and making the lasers very large is impractical. The solution to this problem is to contain the active material (where the gain is generated) within a laser cavity. The cavity is a structure which causes a proportion of the photons to feed back into the active region rather than being emitted, increasing the effective length of the active material. The simplest and most common example is the Fabry-Perot cavity, which consists of a partially reflective surface (mirror) at each end of the active region in a plane-parallel arrangement as shown in figure 2.11.

![Figure 2.11: A Fabry-Perot cavity consisting of two mirrors in a plane-parallel arrangement around a laser active region.](image)

The mirrors can have a differing reflectivity, one much greater than the other to improve the cavity Q-factor (the fraction of energy stored to lost from a resonator) since generally emission is only desired from a single facet. In the case of a semiconductor laser the mirrors commonly consist only of cleaved facets that have a reflectance of $\sim 25\% - 30\%$ resulting from the refractive index ($\sqrt{\epsilon_r}$, where $\epsilon_r$ is the materials...
relative permittivity) difference between the semiconductor and surroundings (generally air). The reflectance is defined as

\[ R = \left| \frac{\sqrt{\varepsilon_{r1}} - \sqrt{\varepsilon_{r2}}}{\sqrt{\varepsilon_{r1}} + \sqrt{\varepsilon_{r2}}} \right|^2, \]  

where \( \sqrt{\varepsilon_{r1}} \) and \( \sqrt{\varepsilon_{r2}} \) are the surrounding and semiconductor refractive indices with values of approximately 1 and 3 respectively. If the as cleaved facets are found to be insufficiently reflective a range of high reflectivity coatings can be deposited on the surface to improve the Q-factor. In addition to confinement of photons by reflection, the cavity forms standing wave longitudinal modes, acting as a resonator and limiting the wavelengths that may propagate, to those for which an integer number of half wavelengths will fit into the cavity. If the wavelength/frequency/energy of adjacent cavity modes are calculated and subtracted the mode spacing can be calculated, giving the following relation:

\[ \Delta \nu = \frac{c}{2\sqrt{\varepsilon_r}L}. \]  

It can be seen from 2.26 that the shorter the cavity length, the larger the frequency spacing, \( \Delta \nu \), causing the laser to become increasingly wavelength selective. Thus small devices, such as Vertical Cavity Surface Emitting Lasers (VCSELs) will have large cavity mode spacings, giving the possibility for a very selective emission energy. The transmission spectrum of the cavity modes and the gain spectrum above transparency are plotted together in the top part of figure 2.12. Whilst the gain spectrum is a continuous function of energy, the cavity modes quantize the photon energies that may propagate. Combining these two functions allows the emission spectrum to be determined, as is shown in the bottom plot of 2.12. The mode spacings are shown for a fairly standard 1 mm cavity length, where the mode spacing is only \( \sim 0.2 \) meV, giving an almost continuous set of modes. To clarify the discussion the modes for a 30 \( \mu m \) cavity length are shown since the spacing and effect of the modes is more visible.

Above transparency several modes will be emitted simultaneously, however each experiences different levels of gain. The state described by figure 2.12 is above transparency but not yet at lasing threshold. As threshold is approached the gain increases and therefore the difference in photon density of each mode also becomes greater. Modes towards the edge of the gain spectrum will therefore start to become negligible as carriers are de-populated at an increased rate by the modes closer to the gain peak. This leads to a narrowing of the emission as lasing threshold is approached and modes towards the centre are favoured, with only the few modes of greatest gain (or a single mode in some cases) remaining above threshold. Further details of the lasing threshold are found in section 2.6.3.
Figure 2.12: Top: The gain spectrum above transparency and transmission in arbitrary units for a hypothetical device. Bottom: The product of the gain and transmission, giving the form of the emission spectrum.

2.6.2 The distributed feedback cavity (DFB)

An alternative to the Fabry-Perot cavity is to form what is known as a distributed feedback or DFB cavity. Rather than having a mirror at each end of the resonator a periodic structure is defined throughout the active region or waveguide, somewhat akin to a diffraction grating to give feedback throughout. The periodic structure is designed so as to meet the Bragg condition (equation 2.27), which for normal incidence reduces to

\[ d = \frac{j\lambda}{2}, \]

\[ 2d \sin(\theta) = j\lambda. \]  

(2.27)

Therefore if the DFB structure is chosen to match the desired wavelength of operation for the laser then this mode will be reflected back into the resonator throughout the structure preferentially. This allows lasers with such a cavity to approximately operate at a single mode, rather than hopping between modes or multimoding, as can take place in a Fabry-Perot cavity due to its lesser selectivity and dependency on cavity length. Spatial hole burning, the process of gain saturation at anti-nodes of the standing wave can allow alternative modes to propagate, increasing the linewidth of the laser. In fact for a totally symmetric DFB the counter-propagating modes at the Bragg wavelength will destructively interfere, causing two modes either side of Bragg wavelength to propagate. However these effects can be somewhat nullified by the introduction
of a quarter wave phase shift within the cavity.

In addition to providing a more spectrally pure emission the DFB waveguide also give a more temperature stable emission wavelength since the refractive index and layer thicknesses are generally less temperature sensitive than the peak gain energy. This is in contrast to the Fabry-Perot cavity which will always lase at the cavity mode closest to the gain peak.

The fact that a only a single axial resonator mode propagates can also increase the sub threshold efficiency and reduce the threshold carrier density since stimulated emission at other modes within the gain spectrum is suppressed. This is however, dependent on the gain-cavity-mode alignment.

The DFB cavity can be particularly useful for integrated systems since facets are not required, allowing the whole laser to be buried within another material, such as Si, for which the index contrast would be too low to form a Fabry-Perot cavity. Furthermore several devices with different pitches (periodicities) can be used to produce several slightly different, spectrally pure emission wavelengths, allowing for the use of wavelength multiplexed signals, delivering a greater density of data for a single waveguide.

2.6.3 Threshold gain and threshold carrier density

The Bernard-Duraourg condition assumes that the only losses of photon density are as a result of absorption and represents the minimum pumping theoretically possible to achieve lasing. In reality there are several other sources of loss that lead to the requirement of additional pumping beyond transparency to achieve lasing. The primary additional source of loss is through at least one of the facets, allowing for the lasing light to be coupled out. But additional sources such as inter valence band absorption, free carrier absorption, scattering and lateral spreading of the photons, among others, will contribute to a varying extent. The effect of these additional sources of loss is to require greater gain in order to counteract their reduction of photon density, taking these into account gives equation 2.28:

$$g_{th} = \frac{1}{\Gamma} \left[ \alpha_i + \alpha_m \right].$$  \hspace{1cm} (2.28)

Where $g_{th}$ is the threshold gain, $\Gamma$ is the optical confinement factor (the fractional overlap between the optical mode and the active region), $\alpha_i$ represents the internal losses per unit length (scattering etc) and $\alpha_m$ represents the mirror losses per unit length. $\alpha_m$ is defined by equation 2.29, where $R_1$ and $R_2$ are the reflectivities of the two cavity mirrors/facets and $L$ is the cavity length (not necessarily the same as the active region length):

$$\alpha_m = \frac{1}{2L} \ln \left( \frac{1}{R_1 R_2} \right).$$  \hspace{1cm} (2.29)

The confinement factor is calculated by taking the integral of the optical field intensity in the QWs and dividing by the integral of the optical field over its full extent in space [155]. This is defined by equation 2.30,
where $z$ is the growth direction. The modal gain, $\Gamma_g$ takes into account the incomplete overlap between the optical mode and the QWs, giving the actual gain that can be expected from the active region. Increasing the modal gain through maximizing the confinement factor will cause a superlinear decrease in threshold current density. This is thus an important aspect of device optimization, as shall be discussed in chapter 6.

From 2.28 and 2.29 it can be seen that the condition for lasing is for the gain to be equal to the total losses when a photon completes a round trip journey within the cavity, i.e. for the net losses to be zero for a round trip journey.

The carrier density is related to the gain by the empirical McIlroy relation [116] (equation 2.31),

$$g = g_0 \ln \left( \frac{n}{n_{tr}} \right),$$

(2.31)

where $g_0$ is the material dependent gain coefficient and $n_{tr}$ is the carrier density at transparency, i.e. when $g = 1$. At threshold, the condition $g = g_{th}$ is met, and so substituting 2.28 into 2.31 and rearranging for $n$, which will be equal to $n_{th}$ at threshold gives equation 2.32,

$$n_{th} = n_{tr} \exp \left( \frac{1}{\Gamma g_0} (\alpha_i + \alpha_m) \right).$$

(2.32)

This equation then gives the carrier density that will be required to reach lasing threshold, which in an ideal laser will be pinned at the threshold value (discussed below) regardless of how much additional pumping power is applied once steady state operation has been reached. The reason for this pinning is known as gain clamping, if momentarily $n > n_{th}$ then $r_{stim}$ will increase and as a result so will $P(h\nu)$, further increasing $r_{stim}$. This will pull $n$ back down towards $n_{th}$ and so the effect of adding additional carriers through harder pumping serves only to increase the photon density with the average carrier density staying constant. Similarly if the photon density becomes momentarily too large the carrier density will drop below the steady state threshold value, with $r_{stim}$ falling accordingly until the carrier density recovers to the threshold value. These fluctuations are known as relaxation oscillations and are depicted in figure 2.13.

Thus it can be seen that adding additional carriers through further pumping serves only to maintain the steady state carrier density at the threshold value as the stimulated emission rate rises, therefore in an ideal case leading to all additional carriers being converted to stimulated emission beyond threshold. Equation 2.21 shows that the various recombination mechanisms have a differing dependence on carrier density and so the fractional contribution of each will vary with $n$, the pinning of $n_{th}$ means that above threshold the fractions of each recombination will also become fixed if all other factors remain equal. This is of particular importance for an Auger or carrier leakage dominated device since these mechanisms scale so strongly with carrier density. In practice other factors can cause the carrier density not to pin, as discussed later.
2.6.4 Free carrier absorption (FCA) and inter-valance band absorption (IVBA)

Two of the main additional optical losses that lead to the threshold gain being larger than the transparency point (other than mirror losses) are free carrier absorption (FCA) and inter-valence band absorption (IVBA). These two processes represent the case of optical absorption that take place between energy levels other than those associated with the lasing transition. Free carrier absorption represents the case of a free carrier (electron in the conduction band or hole in the valence band) absorbing a lasing energy photon and undergoing an intraband transition. Above threshold the stimulated emission rate is ideally so great that all injected carriers rapidly recombine via stimulated emission and therefore the FCA cross section is so low as to make the process negligible.

Therefore this process is much more prominent below threshold and outside of the active region wherever the optical mode overlaps with an injection region. For example, a barrier layer (see figure 2.7) will have a large electron or hole density during pumping depending if it is on the n or p-side of the junction; meaning that there are free carriers that could absorb photons emitted from the junction if they spread into these regions. This process can become particularly prominent in doped regions since the carrier density is always so much larger, even when carriers are not being pumped through the layer.

Inter-valence band absorption represents the case of photons emitted from band-to-band transitions being absorbed by electrons deep enough into the SO band to be excited into the HH or LH bands. Unlike Auger recombination there is no additional momentum carrying body to balance non-vertical transitions within the band structure. Thus, there must exist somewhere in the valence band states for which there are both

Figure 2.13: The relaxation oscillations of the carriers and photons as threshold is reached.
electrons in the SO band (generally a given) and holes in a LH or HH state at the same wave vector. The valence band quasi-Fermi-level does not even always reach the valence band edge at threshold due to the high mass of holes in comparison to electrons (the electron Fermi level extends further into the conduction band to compensate) so generally only the uppermost states of the valence band are filled. Therefore the rate of IVBA increases as the SO splitting tends towards the energy of the lasing transition energy so that IVBA may take place near to $k = 0$ where the hole density is greatest. Being dependent on hole density this process becomes much more likely within p-doped material.

Both of these optical loss processes serve to increase the internal loss of the device per unit length, $\alpha_i$, increasing the gain required to reach threshold and therefore also the threshold carrier density and threshold current density, the efficiency of the device above threshold is also reduced [6] as photons that could have contributed to the emission power are lost to these processes.

2.7 Threshold current and slope efficiency

The threshold carrier density is not easily directly measured, and so lasers are generally defined according to a threshold current or threshold current density (threshold current divided by the pumped cross-sectional area of the laser). Quantification of the threshold current is easily achieved by plotting a light-current characteristic (LI) which tracks the emission from the laser facet as a function of injection current (pumping). Since at threshold, the efficiency of photon generation suddenly increases by orders of magnitude once stimulated emission dominates radiative recombination, a sudden change in gradient of the LI is observed. This can be seen in figure 2.14 for which the threshold current is 50 arbitrary units on this scale. Above threshold the gradient of the LI plot gives $\eta_d$, the differential quantum efficiency, slope efficiency or external quantum efficiency, which if measured absolutely gives the efficiency with which injected charge carriers are converted to photons emitted from the facets of the laser per unit time above threshold:

$$\eta_d = \frac{e}{h\nu} \frac{dL}{dI}.$$
It should be noted that the spontaneous emission also exhibits a discontinuity in gradient at threshold, again shown in figure 2.14 (purple data points). This is a result of the carrier density pinning at threshold and therefore causing equation 2.9 to become constant as with injection current. Deviation from this behaviour is a clear identifier of significant issues with the laser and gives cause for more in-depth investigation to determine the exact cause.

2.8 Internal quantum efficiency

Equation 2.33 relates the external quantum efficiency to easily measurable parameters and describes the differential efficiency of the device above threshold. Although useful it is also desirable to determine the efficiency with which the carriers are converted into photons within the device, isolating the efficiency from optical losses. To do so, the internal quantum efficiency, $\eta_i$ is defined as follows [12]:

$$\eta_d = \eta_i \left( \frac{\alpha_m}{\alpha_i + \alpha_m} \right).$$

The internal quantum efficiency below threshold is defined as:

$$\eta_i(n < n_{th}) = \frac{\tau_{nr}}{\tau_r + \tau_{nr}}.$$  \hspace{2cm} (2.35)

$\tau_r$ and $\tau_{nr}$ are the radiative and non-radiative lifetimes of carriers within the active region, defining the
fraction of recombination that is radiative (spontaneous or stimulated emission). Above threshold however the definition changes somewhat to take into account that only lasing photons, produced by stimulated emission are useful, with spontaneous emission now a source of inefficiency:

\[ \eta_i(n = n_{th}) = \frac{I_{stim}}{I - I_{th}}. \]  

(2.36)

\( I_{stim} \) is the current associated with stimulated emission, \( I \) is the injection current and \( I_{th} \) is the threshold injection current. Ideally \( I - I_{th} = I_{stim} \) since all carriers injected beyond threshold contributed to stimulated emission and lasing, giving \( \eta_i = 1 \) above threshold. However, in reality factors such as imperfect pinning can cause additional non-radiative recombination to take place or internal heating can degrade performance. If 2.34 is rearranged it can be written as:

\[ \frac{1}{\eta_d} = \frac{1}{\eta_i} \left( \frac{\alpha_i}{\alpha_m} + 1 \right), \]  

(2.37)

which when combined with equation 2.29, the definition of \( \alpha_m \) the following relation is yielded:

\[ \frac{1}{\eta_d} = \frac{1}{\eta_i} \left[ \frac{2\alpha_i L}{ln \left( \frac{1}{R_1 R_2} \right)} + 1 \right]. \]  

(2.38)

This equation allows both the internal quantum efficiency and internal loss to be related to the practically measurable parameters \( \eta_d \) and \( L \), the cavity length. If \( \eta_d \) is determined for several devices of varying cavity length then by plotting the inverse slope efficiency against cavity length then the y-intercept will be equal to \( \frac{1}{\eta_i} \) and the slope will be proportional to the internal loss per unit length (with \( R_1 \) and \( R_2 \) known from the refractive index contrast between the semiconductor and surrounding material). It should be noted that this treatment is only valid for a Fabry-Perot cavity and if the carrier density pins.

### 2.9 Z analysis

#### 2.9.1 Calculating Z

According to equation 2.21 each recombination mechanism exhibits an individual dependence on the carrier density, \( n \). Shockley-Read-Hall, radiative, and Auger recombination have a linear, quadratic and cubic dependence respectively, with the leakage current dependence given by equations 2.16 - 2.20 (essentially exponential dependence on \( n \)). In order to give an indication of the dominant recombination mechanism, the carrier density dependence of the total current can be quantified with a “\( Z \)” parameter [131]. \( Z \) is calculated by plotting the logarithm of the total current against the logarithm of the square root of the integrated pure spontaneous emission (from equation 2.9 \( L^{1/2} \propto n \)), as is shown in figure 2.15.

Thus the gradient of the plot, \( Z \), will be equal to the carrier density power-law dependence of the total current (on a logarithmic scale, \( \frac{d \ln(I)}{d \ln(n)} \)). Since the lowest value \( Z \) can take is 1, representing pure SRH recombination, extrapolation from a point where \( Z \approx 1 \) to the threshold current will give the proportion of
the total current accounted for by SRH recombination [131]. The $Z$ value at threshold then gives a measure of the carrier density power-law dependence of the threshold current, which can be compared to the expected values for the hypothetical limits of devices dominated entirely by each mechanism (i.e. $Z = 2$ for radiative, $3$ for Auger etc) to indicate which are dominant in the device under investigation.

![Figure 2.15: Example plot for extraction of the “$Z$” parameter from the spontaneous emission as a function of current. The red dotted line marks the threshold spontaneous emission power and the black line is a tangent at $Z = 1$.](image)

2.9.2 Determining recombination mechanism fractions through $Z$ analysis

In cases where carrier leakage can be assumed negligible it is possible to approximate the fractional contribution to the total current from that of each recombination mechanism. The mathematical procedure, starting from the definition of the $Z$ parameter is as follows:

$$Z = \frac{d \ln(I)}{d \ln(n)} = \frac{1}{I} \frac{dI}{dn} = \frac{n}{I} \frac{dI}{dn}.$$

Equation 2.21 gives the breakdown of injection current as a function for each recombination mechanism as follows (assuming negligible carrier leakage):

$$I = eV(An + Bn^2 + Cn^3).$$

Therefore the derivative with respect to $n$ will be:

$$\frac{dI}{dn} = eV(A + 2Bn + 3Cn^2).$$
Combining these with the definition of $Z$ above then gives:

$$\frac{n}{I} \frac{dI}{dn} = \frac{eV}{I} \left( An + 2Bn^2 + 3Cn^3 \right) = \frac{eV}{I} \left( An + Bn^2 + Cn^3 \right) + \frac{eV}{I} \left( Bn^2 + 2Cn^3 \right),$$

which reduces to:

$$\frac{n}{I} \frac{dI}{dn} = \frac{I}{I} + \frac{I_{rad}}{I} + \frac{2I_{Aug}}{I} = 1 + \frac{I_{rad}}{I} + \frac{2I_{Aug}}{I}.$$ 

Using this process a definition of $Z$ that consists of a combination of defect, radiative and Auger recombination fractions is formed, which if weighted correctly, can be combined to fit specific $Z$ values to determine what combinations of recombination mechanisms are plausible.

### 2.10 Characteristic temperatures, $T_0$ and $T_1$

Temperature variations can and do cause changes in the various recombination rates and to different degrees for each mechanism. The cause can be process specific, such as the Auger activation energy or more generally related to the temperature dependence of the carrier density inherited from the Fermi-Dirac distribution. These rate changes will have the knock on effect of altering the associated currents described in sections 2.2 and 2.3. The impact of this effect is to change the threshold current of the laser with temperature. Thus, by monitoring the threshold current as a function of temperature, information pertaining to the dominant recombination mechanisms can be extracted. The characteristic temperature is a quantitative measure of the temperature sensitivity of the threshold current and is given by:

$$T_0 = \left[ \frac{d \ln(I_{th})}{dT} \right]^{-1} = \left( \frac{1}{I_{th}} \frac{dI_{th}}{dT} \right)^{-1}. \quad (2.39)$$

If $T_0$ is constant and we define $I_0$ as the threshold current at zero Kelvin, then we can state that:

$$I_{th} = I_0 \exp \left[ \frac{T}{T_0} \right]. \quad (2.40)$$

Thus from equation 2.40 it can be seen that a high $T_0$ is desirable since this corresponds to a lower temperature sensitivity of $I_{th}$. Given that the temperature dependencies of the recombination mechanisms differ, we can define the characteristic temperature for lasers dominated by each mechanism as follows: $T_0 = \frac{T_0}{2}$ represents a defect dominated, $T_0 = T$ a radiatively dominated, $T_0 = \frac{T}{3} + \frac{T_{Aug}}{}$ an Auger dominated and $T_0 = \frac{T}{(k_B/k_B)}$ a carrier leakage dominated quantum well laser. Derivations of these values can be found in section 2.5 of [58]. Similarly to $T_0$, $T_1$ quantifies the temperature sensitivity of the differential quantum efficiency and is defined as follows:

$$T_1 = - \left[ \frac{d \ln(\eta_d)}{dT} \right]^{-1}. \quad (2.41)$$

A large $T_1$ indicates a low temperature sensitivity and therefore temperature stable device, desirable for many
applications such as integration with electronic components where temperatures may change with system load for example. Values of $T_1 \approx 1000$ K are not unusual in commercial grade devices [59] but can drop as low as single digits in particularly sensitive devices that are limited by optical losses or a degrading internal efficiency.

2.11 Hydrostatic pressure

2.11.1 Influence on the crystal and band structure

Hydrostatic pressure refers to the pressure applied by a fluid in a confined space. If confined, the fluid will exert a force equally in all directions on its surroundings. Therefore if a semiconductor sample is placed with a confined space and a pressurized fluid introduced then it will experience an equal force acting on it from all directions. The effect of this is to compress the crystal lattice uniformly, therefore maintaining the symmetry of the crystal but reducing the lattice constant. Since the band structure is tied to the lattice constant this will also be varied, with an example shown in figure 2.16 for $Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$, which is investigated as part of chapter 7. Each band is characterised by an individual pressure coefficient, quantifying the sign and extent of the change per unit pressure, generally measured in kbar or MPa. The general trend is for the $\Gamma$ and L-minima conduction band energies to increase with pressure whilst the X-minimum of the conduction band moves down with pressure, with the coefficients varying with material.

If the pressure coefficients for binary compounds are known then interpolation of these is found to accurately reflect the coefficients of ternaries and quarternaries [102]. The relative alignment of the $\Gamma$, X and L-minima can therefore be varied by the application of hydrostatic pressure, which may be impactful in devices for which there is a reasonable probability of carriers leaking into the satellite bands. In addition to the bands within a given material the alignment across junctions between dissimilar layers will also vary as a result of the material dependency of the pressure coefficients.
2.11.2 Semiconductor laser characterisation as a function of hydrostatic pressure

Since the various recombination mechanisms are each dependent on either the band gap or band alignments to varying degrees varying these parameters can be used as a method of characterising these processes. Generally this is achieved by growing a series of structures each with a slightly different composition, however doing so is not only both time consuming and expensive but also introduces an uncertainty in the results associated with differences in growth conditions that may exist between batches. Furthermore changing only the band alignment or band gaps without effecting any other parameters, such as strain or layer thickness is not practical, adding yet more uncertainty to results based on this approach. The application of hydrostatic pressure however does not suffer from these short comings, since a single device can have its band alignments varied to varying degrees in a reversible manner.

The radiative component of the injection current in the limit of a defect free ideal laser that is not limited by non-radiative recombination is found to be approximately proportional to the square of the bandgap [5]:

\[ I_{\text{rad}} \propto E_g^2. \]  

(2.42)

With reference to figure 2.16 the implication is that the radiative current as a function of pressure will increase at a rate proportional to \( (E_g(0) + \frac{dE_g}{dP} P)^2 \), where \( E_g(0) \) is the ambient pressure band gap and \( \frac{dE_g}{dP} \) is the \( \Gamma \)-minimum pressure coefficient. In the limit of a defect dominated device one would expect a pressure independent defect current with pressure, since the recombination rate via deep level defects should be independent of the band gap or alignments. In the case of Auger recombination the trend with pressure is generally a decrease with pressure resulting from an increased \( E_g - \Delta_{SO} \) splitting on which the CHSH process...
is dependent (equation 2.14) or just $E_g$ in the case of CHCC (equation 2.15).

The pressure dependence of $J_{\text{leak}}$ is a combined effect of the pressure coefficients of the level from which the carriers originate (the Fermi-level of the active region at the $\Gamma$-minimum) and the level into which the carriers leak (a level within the barriers for example). The offset between these two levels is the effective confining potential, $E_b$ defined in section 2.3.3 [117]. The pressure coefficient of the confining potential is determined practically by measuring the pressure dependence of lasing emission and adding to this the pressure dependence of the leakage level, generally calculated through interpolation of the binaries. $J_{\text{leak}}(0)$ is the leakage current density at ambient pressure, which can be used with equation 2.43,

$$J_{\text{leak}}(P) = J_{\text{leak}}(0) \cdot \exp \left( -\frac{dE_b}{dP} \frac{P}{k_bT} \right) ,$$

(2.43)

to determine the trend in leakage current density with pressure [34]. In the case of a carrier leakage dominated device the pressure dependence of $J_{\text{th}}$ can be determined as follows [117]:

$$\frac{J_{\text{th}}(P)}{J_{\text{th}}(0)} = \left( \frac{J_{\text{rad}}(0)}{J_{\text{th}}(0)} \cdot \frac{J_{\text{rad}}(P)}{J_{\text{rad}}(0)} \right) + \left( 1 - \frac{J_{\text{rad}}(0)}{J_{\text{th}}(0)} \right) \cdot \frac{J_{\text{leak}}(P)}{J_{\text{leak}}(0)} .$$

(2.44)

The pressure dependencies of these mechanisms are plotted in figure 2.17 for the case of a GaSb based device investigated in chapter 7. These limiting dependencies can be used as to quantify the contributions of each recombination mechanism when compared to the threshold current as a function of pressure (see section 7.2 of chapter 7).

**Figure 2.17**: Illustration of the pressure dependencies of the defect, radiative, Auger and leakage currents.
2.12 Performance improvements

So far, the most basic laser design of a Fabry-Perot cavity homojunction laser has been assumed for simplicity of discussing the fundamental properties of the devices. In reality however there are a variety of improvements beyond the homojunction that are in common use within the majority of semiconductor lasers. In this section, the use of quantum well active regions and separate confinement heterostructures are presented since these are relevant to the devices discussed in this thesis. Figure 2.18 is a schematic diagram of a separate confinement heterostructure (SCH) which consists of a quantum well active region surrounded by barriers for carrier confinement and cladding layers for optical confinement. These concepts and their impact on the devices is discussed in the subsections that follow.

![Figure 2.18: Top: Carrier confinement using a QW structure (CB shown). Bottom: Optical confinement by index guiding using a SCH structure.](image)

2.12.1 Quantum confinement

The bulk energy levels given by the dispersion relation (equation 2.1) and density of states given by equation 2.3 lead to a situation whereby many states of differing energy must be filled in order to meet the Bernard-Duraffourg condition. The occupation for bulk semiconductor material (the product of the Fermi-Dirac distribution (equation 2.4) and bulk density of states is shown in the left image of figure 2.19. Basic quantum mechanics states that confining electrons to a region of less than their de-Broglie wavelength causes quantisation of the allowed energy levels which the electrons may occupy, given by 2.45:

$$E_i = \frac{w^2 \hbar^2}{8m^* L^2},$$

(2.45)

where \(w\) is the energy level index and \(L\) is the thickness of the confined layer for an infinite square well. In reality the well has a finite depth but the solution for a finite well is not analytically solvable and so
the approximation of an infinite well, which will give slightly larger energies is given here to highlight the general effect of confinement. Confinement in one dimension, as is the case with a square well, is achieved by layering a sub de-Broglie wavelength thickness (generally $\leq 10$ nm) semiconductor between two layers of larger wavelength known as barriers. This arrangement defines what is known as a quantum well, and is the type of active region that is used in the majority of commercial semiconductor lasers. In addition to quantising the energy levels (marked in red on figure 2.19, right) the density of states is also altered causing it to change from a density of states $\sqrt{E}$ dependence to a Heaviside function that is constant in between steps which coincide with the quantum well energy levels, as given by:

$$\rho_{QW} = \frac{4\pi m^*}{h^2 L_w} w.$$  \hspace{1cm} (2.46)

Multiplying the density of states by the Fermi-Dirac distribution gives the occupancy as a function of energy, shown for bulk semiconductor and quantum well on the left and right respectively of figure 2.19. It can be seen that far fewer states are filled for a given Fermi-energy in the case of a quantum well, increasing the proportion that take place in gain generation, increasing the internal quantum efficiency, lowering the threshold carrier density and threshold current density.

Figure 2.19: The occupation (product of Fermi-Dirac distribution and density of states) for bulk (unconfined) and a quantum well (1D confinement), left and right figures respectively.

In addition to the change in density of states, the quantum well structure confines the carriers within the active region as a result of the band offset between the QW $E_1$ or $HH_1$ states (from which lasing takes place) and the barrier conduction or valence bands. The confinement of electrons in the conduction band of a QW structure is shown in the upper diagram of figure 2.18. Since the electrons are injected from the band edge of the larger band gap materials they are transported to the lower energy gap QW. The QW is therefore far superior in minimising carrier leakage if appropriate barrier materials are chosen, with a band offset $\gg k_b T$. 

43
Adding further QWs separated by thin barriers is known as a multi-quantum well (MQW) laser structure and has the benefit of multiplying the density of states at a given energy by the number of wells. Splitting the carriers between the wells reduces the Fermi-level splitting required to achieve an inversion. Filling each band to a lesser degree leads to an overall lower threshold current density since fewer states are filled that do not contribute to gain and therefore the threshold current density per QW drops. The lower threshold carrier density has the benefit of reducing highly carrier density dependent loss mechanisms such as Auger recombination or carrier leakage, as does the reduction in band filling. Furthermore the increased QW volume leads to gains in optical confinement factor, increasing modal gain and lowering optical losses such as IVBA.

2.12.2 Optical confinement

In order to minimise threshold current density \( J_{th} \) the modal gain, \( g \Gamma \) must be maximised since this will minimise the threshold carrier density \( n_{th} \) as shown by equation 2.32. The optical confinement factor, \( \Gamma \), is the ratio of the optical mode overlap with the active region, as shown by equation 2.30. Therefore in order to maximise this parameter the laser should be designed in such a manner as to act like a wave-guide, confining the greatest proportion of the optical mode to the active region. Guiding the optical mode in the growth direction is achieved by creating as large a refractive index step as possible between the active region and the surrounding layers. Since the refractive index scales inversely with the band gap, a small band gap material surrounded by a large band gap material is desirable. This represents the situation of of a QW, however the well itself is very thin, and as a result provides fairly weak optical confinement even though the index step is reasonable large. Therefore an additional pair of layers either side of the barriers known as cladding, with even larger band gaps and therefore lower refractive index is included creating what is known as a separate confinement heterostructure (SCH). This arrangement is depicted in the lower plot of figure 2.18 and illustrates that for appropriate choices of material composition the SCH structure can act as a reasonable waveguide and confine the majority of the mode to the active region.

In the transverse direction index guiding, although possible and implemented in some cases, such as buried heterostructures is more difficult to implement, since unlike index guiding in the growth direction it is not simply a case of growing differing planar layers. The simplest route of achieving index guiding is to form a mesa structure with the sides exposed to air, giving a strong index contrast, but can be impractical for some applications. As a result, in many broad area lasers, such as those investigated in chapters 5, 6 and 7 the optical mode is confined in the transverse direction by gain guiding. Gain guiding involves amplifying the centre of the mode more than the tails. In its simplest form this is achieved by limiting the width of the contacts so that injection only takes place to the desired width, with a drop off in pumping in the transverse direction from current spreading. The refractive index has been shown to vary with electrical field strength (including that of electromagnetic waves) leading to Kerr-induced self lensing of optical beams. However, this is a non-linear optical process and as such the critical power required for the effect to take place is generally far greater than that which would be expected for these relatively low power devices [14].

Further details of optical confinement are presented in section 6.1.1 of chapter 6, which presents a study
of the maximisation of optical confinement for GaNAsP based lasers through passive waveguide simulation.

2.13 Bandgap renormalisation

The band gap, being dependent on the interaction of electrons with the lattice would logically be affected by electric fields which will vary the strength of the interaction. When operating a laser the carrier density injected into the active region can become very large, particularly in un-optimised devices that have large threshold current densities. This large injection of charged particles can have a shielding effect on the other electrons in the system, reducing the interaction strength between the electrons and the lattice in a somewhat analogous manner to large atoms with many filled orbitals. Therefore whilst negligible at low injection at large current densities the band gap can be reduced slightly by the presence of so many additional free carriers injected into the system, lowering the emission energy, this process is known as bandgap renormalisation [153]. Bandgap renormalisation acts against the Burstein-Moss shift, resulting from band filling, and so can be difficult to directly measure using the emission energy of the laser, since this will reflect the net effect of both processes.
Chapter 3

3 Experimental Processes

This chapter describes the experimental methods that are used to characterise the devices studied in this thesis. Schematics of the experimental set ups and details of their components and how the system is operated are presented in addition to any additional information such as short comings or assumptions required. Each processes allows specific information to be extracted, which will be summarised along with each section. Finally, details of the software and procedures used for simulations are also included. The results of investigations produced using these systems are presented in chapters 4 to 7.

3.1 The general electroluminescence characterisation arrangement

All electroluminescence experiments regardless of the conditions applied to the device under investigation have an identical core collection of standard equipment. Some pieces such as specific detectors or power supplies may be exchanged for equivalent alternatives but the system as a whole is unchanged, this basic arrangement is shown schematically in figure 3.1.

With reference to figure 3.1 the device under investigation, (A) is mounted in a chosen system, (B) (probe station, high pressure system or close cycle cryostat) relevant to the experiment to be carried out and mounted with the appropriate clips, adhesives and electrical contact probes. Details of each specific system and mounting types used for each are given in sections 3.2.1, 3.2.3 and 3.2.2 for the probe station, closed cycle cryostat and high pressure system respectively. A 47 Ω resistor, (C) is connected in series with the laser diode in order to impedance match the coaxial cables, (J) since the devices tend to have resistances of ∼3 Ω. The choice of power supply (D) is based on the specifics of the experiment, but wherever possible a pulsed voltage source is used in order to minimise Ohmic internal heating by lowering the duty cycle. Two types of pulsed voltage source are used, the Avtech 10011B1 and BNC 6040+202H with the choice of which dependent on the injection levels required for the specific investigation. The BNC is able to supply a pulsed voltage up 300 V, compared to only 100 V in the case of the Avtech and so must be used in cases where large biases are required to achieve lasing. The BNC however has a minimum voltage limit of 2 V, which is greater than required to reach threshold in some devices and especially so if investigation of electrical turn on is being carried out. In both cases the pulse shape is found to be close to the ideal square shape, allowing the duty cycle (DC) to be calculated accurately using the following relation:

\[ DC = W f = \frac{W}{t}, \]  

(3.1)

W is the pulse width or “on” time for each square pulse, f is the frequency of pulses (one full on and off state) and t is the time taken for each pulse (1/f). The typical values used are a pulse width of 500 ns and a frequency of 10 kHz, this gives a duty cycle of 0.005 or 0.5%, which has been found to be sufficiently...
low to make internal heating effects negligible whilst producing sufficient optical power for detection with a reasonable SNR. In cases where additional current is required, such as the case of a device with a particularly weak emission or if precise, low bias IVs are required (where the pulse shape can be poor) a CW current source, such as the Keithley 2400 SMU is used instead. The advantage of the SMU (source measure unit) is that it can both supply current and measure the IV to a high precision (~pA) simultaneously, of particular use for electrical investigation.

If an SMU is not used then the voltage are current applied to the laser diode are measured using a Tektronix TDS3052 oscilloscope (E) connected to a current probe (F) and triggered from the TTL output of the power supply. The emission from the laser diode is then coupled into a detector (G), either directly into an integrating sphere if possible to maximise collection efficiency or via cleaved and polished bare fibre if not. The integrating spheres used in these investigations are the InGaAs ILX Lightwave OMH-6708B and ILX Lightwave OMH-6727B, these are easiest to use with the probe station since they can be placed close to the device under investigation but may be used via a window in the cryostat or pressure cell, focused by an appropriate choice of lens. In cases where fibre coupling is required (generally from within the cryostat

---

Figure 3.1: The core arrangement of equipment used for electroluminescence characterisation experiments, the equipment is identified as follows: (A) The laser device under investigation. (B) The specific system in which the device is mounted (probe station, closed cycle cryostat or hydrostatic pressure cell), details of each are given in sections 3.2.1, 3.2.3 and 3.2.2. (C) A 47 Ω resistor. (D) Power supply (pulsed or CW sourcing either Voltage or current based on specific equipment choice). (E) Tektronix TDS3052 oscilloscope. (F) Current probe (G) Detector or optical spectrum analyser (OSA). (H) Temperature controller. (I) GPIB IEEE488 interface bus. (J) Coaxial cable. (K) Computer running LABVIEW 10.1. (figure modified version of original from [58])
or pressure cell) an Ando AQ2140 InGaAs detector is used as it covers the wavelength range of interest and will pick up signals of a few pW, of particular use for inefficient lasers which emit little light. If spectra are to be collected a fibre coupled Ando AQ6315A optical spectrum analyser (OSA) is used, which can resolve weak spectra, measuring only a few 10s of pW of integrated power.

If the temperature of the system is to be varied an OXFORD ITC502S temperature controller (H) is used to control the current applied to a resistive heating element or thermoelectric-cooler (TEC) in order to reach a user defined temperature. If the closed cycle cryostat is used then the cooling is applied manually at all times and the heating element current varied by the temperature controller to achieve a given temperature. For small changes in temperature (10 - 20 K) the system if left for 15 - 20 minutes to reach thermal equilibrium before any measurements are made.

Where possible all equipment is connected to a computer (K) by National Instruments GPIB IEEE488 cable (I) so that the LabView software may be used to automate the systems and to collect and plot data.

3.2 Device mountings

Depending on the contact geometry of the device under investigation and the system being used to do so different mounting systems are used. The mounting serves several purposes: Firstly the device must be held firmly in place in order to maintain a constant alignment with detection systems and therefore give a consistent collection efficiency. Secondly the mounting forms the electrical contacts from the characterisation system to the device so that electrical pumping may be applied. Thirdly the mounting acts as a heat sink to either allow heat to escape the device (particularly important under CW injection scenarios) or to allow active heating or cooling to a target temperature via the cryostat cold finger (see section 3.3) or probe station TEC. Finally some, although not all of the mountings allow spontaneous emission to be collected from the substrate of the device via specifically designed fibre holders.

3.2.1 The probe station

The probe station consists of a large brass plate (for good electrical and thermal conductivity) on an XYZ stage onto which the device is placed. A variant of this plate has a fibre bush mounting for the collection of spontaneous emission, the details of which are identical to that discussed in section 3.2.2. Around the brass plate is a steel shelf on which a number of magnetic probes are placed, which again have XYZ controls. These probes can have a variety of tip sizes fitted which can be precisely aligned with the contacts of the device to allow electrical injection. Below the plate a TEC is placed and used to vary the plate temperature between 10 °C and 60 °C. This system is used where low temperature or high pressure is not required or not possible, because of its open nature there are few constraints on what combinations of equipment may be used in conjunction with this system.
3.2.2 The single contact laser clip

Inside both the closed cycle cryostat and pressure cell space is limited to \( \sim 4 \) and \( \sim 1 \) cm\(^3\) respectively, meaning the mountings must be as compact and reliable as possible since they cannot be adjusted during the course of the experiment. For the case of devices with a contact layout of one on top and one below the laser clip developed by S. J. Sweeney is used [148] and is depicted in figure 3.2.

![Image of single contact laser clip](34)

The clip consists of a brass base with a 5 mm diameter hole into which a brass bush is inserted, this in turn has further hole of 0.5 mm into which the end of a bare optical fibre is secured with resin and polished. If the device is placed on top of this bush then pure spontaneous emission may be collected from the substrate directly by the polished fibre, since gain will be minimal in the vertical direction as the active material thickness will be only that of the QW (\( \sim 5-7 \) nm per well) and there is no cavity defined in this direction.

A thin, top contact, also made of brass is formed in such a way as to act like a spring and when secured at the rear by insulating screws and isolated from the base by a further insulating layer applies a constant force to the top of a device as well as allowing electrical injection. The top contact is raised by depressing a lever to release the spring tension. The second electrical contact is formed from the base plate of the clip, which also acts as the heat sink. Wires from the system of choice are soldered to the appropriate parts of the clip allowing easy removal. Collection of emission from the facet can be either coupled directly into a bare polished fibre end or collected through a window in the cryostat or pressure cell and focused onto a detector using a lens.

3.2.3 The two pin laser mount

For devices mounted in the closed cycle cryostat which feature both p and n-type contacts on the top surface of the laser the single contact laser clip may not be used. In these cases a two pin laser mount is used, which
is shown in figure 3.3. Here the base plate is essentially the same to that of the single contact laser clip and includes the same fibre mounting bush. Rather than a top clip two insulating screws and spacers are used to fix flexible brass probe arms at a user defined angle. To these probe arms probe tips of the type used on the probe station are soldered to the approximate height and orientation for a given sample. The sample is then placed on top of the bush mounted fibre and the pins lifted into place manually, the tension in the flexible probe arms holds the probes and the device in place during experimentation.

![Figure 3.3: The two pin mounting system developed as a modification of the single contact laser clip for device with both contacts on the top surface. The image is edited from original image drawn by Dr. Shirong Jin.](image)

### 3.3 The closed cycle cryostat

For low temperature analysis the device is mounted in either type of laser mount within a cold finger closed cycle helium cryostat, as depicted in figure 3.4. A compressor is used to apply sufficient pressure to helium in order to liquify it, after which it is pumped through high pressure lines into the cold head of the cryostat. The compressed helium is then able to expand in the cold head absorbing heat from the cryostat in doing so, it is then returned to the compressor via a return line and the heat extracted by a water cooled heat sink and the process to repeats. The cold head is attached directly to a cold finger, which acts to extract heat from the sample holder at the end and conduct it to the cold head. The cold finger is held under vacuum conditions to prevent the absorption of atmospheric heat and the build up of ice on cold finger and device. This is achieved by covering the cold finger and device mounting with a steel casing that forms an air tight seal, with the chamber pumped down to a vacuum of \( \sim 10^{-6} \) mBar using both a rotary and turbo molecular pump. The end of the cold finger has a heating coil which is used to counteract the cooling effect of the cold head in order to achieve intermediate temperatures between ambient and a minimum achievable temperature of approximately 20 K, depending on how good a vacuum is achieved. Beyond the heating coil is a copper plate, onto which the device mountings discussed in section 3.2 are attached. A silicon thermocouple is attached to the copper plate allowing the temperature to be monitored by the temperature controller so that appropriate
heating is applied for a given set temperature. Optical fibres pass from the substrate bush and a clip adjacent to the device facet in addition to electrical wires from the mounting contacts back along the cryostat and out through the fixed section of the casing via resin sealed openings. In addition to the fibres light may also be collected from observation windows if sufficiently intense, allowing a wider variety of detectors to be used.

![Figure 3.4: A schematic diagram of the cold finger closed cycle cryostat used for characterisation over the temperature range 300 - 20 K.](image)

Temperature dependence studies have been carried both decreasing the temperature from 300 K to the target minimum value, generally 20 - 40 K as well as back up again in order to check for any hysteresis type effects, changes to alignment or device degradation that may have taken place. By reducing the temperature of a semiconductor laser the crystal contracts, causing the band gap to grow. In addition the average thermal energy of the carriers ($k_bT$) falls, reducing the spread of carriers beyond the Fermi-energy, a factor on which several non-radiative recombination mechanisms and sources of loss are dependent. At lower temperatures the devices should become increasingly efficient with the temperature dependent change in laser characteristics providing an insight into the dominant efficiency limiting processes of the device under investigation. Low temperature characterisation is also a method of producing results from inefficient devices that may not lase at all at ambient temperature. This type of investigation can rapidly yield a great degree of information on the efficiency limitations of laser diodes and so is carried out on all devices investigated in this thesis where possible with results presented in chapters 4, 5 and 7.

### 3.4 High hydrostatic pressure system

The application of hydrostatic pressure to a semiconductor laser will compress it equally in all spatial dimensions (XYZ) causing a uniform reduction in lattice constant. When the pressure is reduced once more the lattice is able to expand back to its ambient pressure equilibrium and since the Γ-point band gap scales inversely with lattice constant it can be altered reversibly using this technique. This process therefore has some similarities to the low temperature characterisation, but differs in the fact that the thermal energy of the carriers is not affected and so changes in recombination fractions will be based on the band gap and band alignment alterations alone. Therefore these two methods are complimentary and so if one carries out
both low temperature and high pressure analysis results that may be somewhat ambiguous using one method or the other may be clarified. The high pressure system was developed to avoid having to grow multiple structures of slightly varying composition, as would be done otherwise to achieve a similar effect. Clearly the reversible changes made to a single structure by the application of hydrostatic pressure has many advantages, most notably a reduction in cost and time. However other issues such as minor differences in growth and processing conditions between the various structures and even across individual wafers are eliminated by using a single sample, making the results more representative of the band alignment changes rather than other external effects.

The pressure coefficients of the band gaps differ with material and so will cause the relative alignments of adjacent bands to change in addition to their absolute band gaps, even for similar materials differing only in component fractions (e.g. Ga$_x$As$_y$P compared to Ga$_z$As$_v$P). The satellite bands of the materials will also vary with pressure although not at the same rate, or even in the same direction as the Γ-minimum, generally the X-minimum band gap will have a negative pressure coefficient, whilst that of the L-minimum will be positive but smaller than that of the Γ-minimum. The pressure coefficients of non-active layers can be determined by tracking the band gap as a function of pressure using techniques such as photoluminescence (PL) or photomodulated reflectance (PR) spectroscopy. Generally though the pressure coefficients for such layers can be calculated using an appropriate interpolation scheme from the pressure coefficients of the constituent binaries. For the active region the energy of the lasing transition determined from the above threshold emission wavelength is easily determined as part of electroluminescence studies. If the pressure coefficients are known then the theoretical dependencies of the various recombination mechanisms on pressure are predictable, allowing a lot of information on which dominate the properties of the devices to be gathered through characterisation as a function of pressure.

The pressure cell into which the device is mounted is shown schematically in figure 3.5. It consists of a thick walled copper-Beryllium cylinder with removable plugs at each end for inserting the device holder and sapphire window, through which emission escapes to be collected. Disposable brass seals are placed on both plugs and tightened using a torque wrench in order to achieve a seal capable of maintaining pressure, whether the seal holds is heavily dependent on the quality and cleanliness of the seals. The only mounting system which is small enough to fit in the pressure cell is the single contact laser clip (section 3.2.2), the two pin mount (section 3.2.3) is too large to fit within the confines of the pressure cell. Devices with a single top and bottom contact are therefore fairly easy to mount, however those which have both contacts on top are somewhat more challenging. The devices discussed in chapter 5 are received as bars which are not easily cleaved due to the fragile Si substrate and so must be mounted whole or cleaved roughly to a section that will fit if too large, leaving the device of interest far from the cleave point. The p-type contact was aligned with the laser clip and the n-type contacts (common to all individual devices on the bar) were shorted to the base plate using silver epoxy. High pressure helium from the compression stage (see figure 3.6 for details) is forced into the cell using a capillary tube above the sample. The use of helium as a pressure medium is advantageous over hydraulic alternatives as there is less interaction between the medium and the device.
in terms of both refractive index contrast change and chemical reactions (found in coated InGaNAS devices [101]). In addition the pressure cell can be placed within a closed cycle cryostat of its own and be operated over the same temperature range as the closed cycle cold finger cryostat described in section 3.3) since the helium will not freeze like a hydraulic fluid would. With any compressed gas there is however the risk of explosive failure of the pressure cell and so the system is held within a protective chamber and sealed when in operation to minimise the risk of damage to users and equipment. Emission is collected from the facet after passing through the sapphire window via a lens and integrating sphere arrangement in order to maximise collection efficiency. As the detection systems are external many different arrangements can be used without space constraints as required for a specific device or investigation.

Figure 3.5: A schematic of the high pressure cell in which devices are mounted. The plugs at either end may be removed to either access the device or change the window (specific to required transmission spectrum). High pressure helium is forced into the cell via a capillary tube mounted in the top to apply pressure to the sample.

The pressure is generated by a three stage hydraulic cylinder system fed from a compressed helium gas cylinder. The lines and open cylinders are first filled with helium from the gas bottle to a target pressure of ~12 MPa and both the bottle and input valve to the pressure system closed. The first stage cylinder is then slowly raised by the hydraulic oil pump to a pressure of at most 70 MPa, forcing the helium out of this cylinder and compressing it into the lines and remaining two stages. The valve to the first cylinder is then closed and the pressure on the first cylinder released by opening a drain valve. The reduction in volume of the remaining stages allows increasing pressure to be generated within the lines and cell without causing the pressure on the hydraulic oil pump to exceed its maximum of 70 MPa. A measurement can be taken at this stage but not when a stage is only partially raised. The process is repeated for the remaining two stages, reaching maximum pressures of 350 MPa and 1000 MPa (limited by the tolerance of the pressure cell, 1500 MPa it technically achievable from the system) for the second and third respectively. The actual pressure achieved will be dependent on leakage from the stages themselves or the the plugs within the
pressure cell, the large pressures cause even small scratches or the presence of contaminants to cause leakage that limits the attainable maximum pressure. Generally around 800 MPa (8 kbar) is the most required for a given investigation and going beyond this value can often be problematic for the reasons stated above. The pressure is increased slowly in order that the temperature of the pressure cell and sample is not raised significantly (isothermal process) and so as not to put undue strain on the system. When a measurement is complete the valves are opened in reverse order slowly to allow the pressure to gradually drop and eventually be released into the atmosphere before, after which the pressure cell may be accessed once more.

![Figure 3.6: The three stage hydraulic system used to pressurise the cell via a gaseous helium pressure medium.](image)

### 3.5 Device processing

The device processing optimisation study presented in section 6.3 of chapter 6 is focused on how the methodologies can be improved and as a result the majority of the details of which processes were used and how are explained at the time. With that in mind this section serves to explain the basic operating procedures for the major items of equipment used and provide links to further information.

#### 3.5.1 Lithography

The majority of the procedures for lithography are detailed in section 6.3.2.1, leaving only the operation of the SUSS MA1006 mask aligner as the major un-explained step. The aim of lithography is to reproduce a pre-designed pattern precisely in a temporary coating known as photo resist. The name of this substance is attributed to its reaction to UV light, which causes it to become more or less soluble in a specific developer solution depending on if it is a positive or negative resist. One a uniform coating of photoresist has been formed on the sample surface exposure to UV light must be minimised by illuminating only in yellow room conditions (UV to blue region of the spectrum filtered from light sources) so that the resist layer remains in
its initial state. To produce a pattern in the resist a mask is used, consisting of a glass slide with the chosen pattern defined by a deposition of a thin metallic film on one side. If a multi-stage process is to take place the mask will have to be carefully aligned with other features, which may be only a few \( \mu m \) across. In order to do so a mask aligner is used, this holds both the mask and sample as well as a UV light for exposure.

In the case of the SUSS MA1006 the sample is placed on a removable stage which holds in in place from below using a vacuum pump. With the sample in place a second holder above the sample stage is loaded with the mask of choice, metal side down and again secured via vacuum. Looking through the in built microscope above the mask the sample stage is raised until the sample is just below the mask and by looking through the transparent areas of the mask the sample is aligned by X and Y translation as well as an in plane rotation micromanipulator. Once in place the sample is brought in to contact with the mask with a range of user specified contact pressures. The UV lamp intensity and exposure time are set and the lamp brought into position automatically above the mask. The illumination with UV light causes the polymer molecules within the photoresist to form parallel bonds to neighboring molecules in the case of the AZ5214E resist used in this study, making the illuminated resist stronger and more chemically stable in the developer. The developer time is chosen based on previous runs and by trial and improvement to be the minimum time required to dissolve the unexposed resist since even when exposed the resist is not entirely insoluble.

Full operating details for the SUSS MA1006 mask aligner can be found in the lab manual available using the following link [92].

3.5.2 Etching

Little equipment is required during the etching step, the acids are mixed on an acids specific wet bench with nitrile gloves and safety glasses offering adequate protection for those used in the studies detailed here. The acids were mixed to the specified ratio, made up to a volume of 100 ml in each case so as to avoid issues of reduced etch rate over time due to breakdown of the etchant but not so much that a spill would be more than an inconvenience. All samples were handled using teflon tweezers to avoid the reaction of acids with metallic items, which may not only damage these items but also introduce contamination.

The depth of features produces by the etchant were determined using a Tenco instruments alphastep 200 surface profiler, which has a stated horizontal and vertical resolution of 40 nm and 5 nm respectively. In practice this was found not to be the case, horizontal distances were generally correct to within the stated error, however vertical steps generally were found to have an uncertainty in the region of 40 - 50 nm. The profiler is operated by placing the sample on an XY\( \Theta \) stage and aligned roughly by eye with the profiler tip. Pressing the “lower tip” button the tip would automatically move down to the sample surface and the raise slightly to allow movement. Using the display the tip could be set above regions of interest using the stage controls and profiling runs of user defined length and resolution carried out in a straight line across the sample. The difference in feature heights could then be measured using the on screen cursors after adjusting for a non-level surface is required. The manual for the alphastep 200 surface profiler can be found using the following link [115].
3.5.3 Metalisation

Metalisation was carried out using three main pieces of equipment, the Edwards E306A evaporator, the JLS loadlocked sputterer and the PPC RTM 2016-M-2G-FC (208) rapid thermal anneal furnace.

3.5.3.1 The Edwards E306A evaporator

The evaporator consists of a large vacuum chamber formed by a bell jar arrangement within which a metal stage with several large electrical contact points is mounted. Between the contacts metallic baskets of other holders can be mounted in which metals to be evaporated are placed. On applying a large current the holders are heated resistively, first melting and then evaporating the metal, the metal vapor then diffuses away from the holder, unperturbed by an atmosphere and is deposited on the first solid object encountered, condensing due to the substantially lower temperature. The sample to be metalised is hung inverted from a mount above the sample holders, which are operated one at a time by rotation of the contacts one at a time to the power supply terminal. Deposition on the sample can be blocked by the use of a “shield” on an arm that may be manually moved between the source and sample. A crystal oscillator measures the approximate deposition thickness at the sample by the change in frequency of oscillation due to the presence of the metal condensed on its surface.

The sample is mounted in its holder using metallic tape and put in place along with the metals and their holders. The bell jar is placed over the sample and the rotary pump activated, once the pressure drops below \(10^{-2}\) mbar the diffusion pump is then also activated and liquid nitrogen added to cool the pump, increasing the rate of pressure drop. The evaporator should be left to pump down until a vacuum of at least \(10^{-5}\) mbar is achieved, although \(10^{-6}\) mbar is preferable if possible, this takes approximately 2 to 3 hours. Once pumped down the first metal can be rotated into position and the thickness sensor zeroed and the metal density entered (for sensor calibration). The current through the metal holder is slowly raised with the shield in place so that any contaminants “burned off” do not contaminate the sample. Once a reading on the sensor starts to appear the shield is removed and the current increased further until the target deposition rate is achieved. Once the desired thickness is achieved the shield is moved back into place and the current slowly reduced back to zero. The process can be repeated for as many metals as required, rotating the holders into place in turn. In order to remove the sample the pumps are turned off in reverse order and the chamber vented once the sensor reads approximately atmospheric pressure. The manual for the Edwards E306A evaporator can be found using the following link [44].

3.5.3.2 The JLS loadlocked sputterer

The sputterer, like the evaporator is also used for the deposition of metals, but does so in a different way. “Targets” made from high purity metals are attached to magnetrons contained within a vacuum chamber, which argon gas may be introduced in to. The magnetrons accelerate Ar ions into the target surface at high velocity causing atoms to be ejected from the surface and onto the sample. The JLS loadlocked sputterer is mainly automated and so requires little user input during operation.

An airlock on the side of the sputter is used to access the sample holder, after the sample is placed on
the holder the airlock is closed and the chamber pumped down to vacuum. The pumps are activated in the correct order automatically and the airlock will not open on the sputter chamber side until a suitably low vacuum has been reached. The sample is automatically moved in on a conveyor and placed on a second holder which has variable height controls. The sample is moved to the appropriate height in the chamber (in this case the height is set to a value of “65”) and sample rotation is switched on (to give a more uniform deposition). The desired source is chosen, causing it to move in to place above the sample and the correct DC power levels chosen for sputtering of that material (optimal values are displayed on the sputterer) and the time required is set. The deposition is somewhat slower than the evaporator with an average of ~2 - 5 Ås⁻¹ compared to 10s of Ås⁻¹ and so a deposition of a few hundred nm usually takes around 20 minutes.

When complete power to the magnetrons is turned off and either the next target required moved in to place or all are sent back to their neutral positions. To remove the sample the instructions for inserting the sample are simply reversed, venting the airlock once the sample has been moved back using the “unload” function. An online manual for the JLS loadlocked sputterer could not be found but the system does feature its own built in electronic manual within the computer interface including all operational and technical details.

3.5.3.3 The PPC RTM 2016-M-2G-FC (208) rapid thermal anneal furnace

The rapid thermal anneal furnace (RTA) is used to aid in the alloying of metals and to diffuse dopants uniformly throughout the capping in order to prevent a large Schottky-barrier forming at the semiconductor/metal interface. The RTA is fairly simple to use, the system is powered on and the chamber opened and the sample placed on the Si wafer holder, which is then re-inserted and locked in place. Pressing the “Pgmtr tables” button opens up a list of programmed steps for the RTA to carry out with the “deviation” button allowing these to be edited. The first step should always be set to purge and step to 2 to process gas, for the uses detailed in this thesis these two are both Ar, the gasses essentially give an inert environment for sample to be heated in to avoid contamination. The next steps require a temperature set point and a segment time (how long to carry that step out for), the first of which will be the temperature for the sample to be raised to with the time determining the rate at which the temperature rises. If the next temperature is set at the same value then the RTA will hold at this temperature for the chosen segment time (note that the sample temperature will take some time to settle at the target value and so hold segment may not give the exact results expected if the segment time is less than ~5s). Additional temperature increases can be added as required, but if not the temperature can be set back to ambient. The segment time for the decreasing temperature should be long (>10 mins, up to 30 mins for 1000 °C) as the cooling is passive and if the program ends before the chamber is cool the heating lamps may become too hot as their cooling fans will be disabled. The final step automatically added is “finish” which powers down the system to a standby state (hence making sure the cooling time is long enough) ready for the sample to be removed.
Chapter 4

4 Efficiency Limiting Processes of Hybrid AlGaInAs Silicon-Evanescent Lasers

4.1 Introduction and background

4.1.1 Hybrid lasers: an approach to optical integration on silicon microelectronics

As was discussed in chapter 1, silicon itself exhibits a number of problematic properties which limit its effectiveness as an efficient medium for the generation of photons, unlike the traditional III-V semiconductor laser/LED materials used in modern optical communications systems. Silicon however is well known to be a particularly good waveguiding material in conjunction with SiO\(_2\) at the standard communications wavelength of 1550 nm, and therefore is exploited in such a role for an abundance of applications. As a result, it is logical to conclude that an efficient optoelectronic integrated circuit could be produced by each device having two separate regions; a III-V section in which traditional, proven active materials based on GaAs, InP, etc generate photons and a silicon/SiO\(_2\) region which forms a waveguide as well as the traditional micro-electronic CMOS components. However, since silicon exhibits such a large lattice constant and thermal expansion coefficient mismatch to the industry standard III-V materials, defect free, monolithic epitaxial growth onto a silicon surface has not yet been achieved to a workable commercial grade.

Thus an alternative approach is to take a pre-patterned silicon-on-insulator (SOI) waveguide along with a standard epitaxially grown III-V structure and use a suitable bonding method to adhere the two together to form a hybrid structure benefiting from the advantages of each individual material system without suffering from the adverse effects of attempting to monolithically grow one upon the other. Several types of bonding have been demonstrated and investigated including; adhesive bonding [165], anodic bonding [98], self-aligned solder bumps [48] and transfer printing [54].

The bonding processes applied in each of these examples although drastically different suffer from a single common weakness. Individual dies must be defined from the III-V wafer, aligned to the silicon substrate with a precision of less than a micron, be that passively or actively and bonded, in some cases individually and others in batches. The surface area of each die for the current Intel processor (Haswell) is stated to be 177 \(mm^2\) in the proceedings of the 2014 ISSCC conference [50], the dies are grown on a 450 \(mm^2\) wafer and so accounting for unusable material around the edge each wafer is expected to yield 770 dies. Each die contains 4 logical cores, GPU, 4 L3 caches, RAM and I/O controllers. Thus even if each die only required one laser the manufacturing of each wafer would be slowed considerably by alignment and bonding of 770 individual laser die, potentially many multiples more. As a result increased manufacturing time alone would raise the cost per unit considerably, since the price point of processors of a given manufacturing run is generally set by the degree to which supply meets the huge, and ever growing demand.

In this chapter the hybrid AlGaInAs silicon-evanescent laser developed by the Bowers group at the
University of California Santa Barbara (UCSB) is studied and characterised. This approach uses a form of flip-chip wafer bonding to directly fuse the AlGaInAs/InP MQW laser material wafer to the pre-patterned silicon on insulator (SOI) distributed feedback (DFB) waveguide. The optical mode propagates within the SOI waveguide and is evanescently coupled into the III-V active region, allowing amplification when biased. Thus, no alignment step is required since the III-V wafer is uniform in the plane of the wafer and whole wafers can be bonded in a single step, rather than individual dies, making the process comparatively fast and consistent with the methods described above when undergoing mass production. A cross-sectional illustration of the lasers investigated in this chapter can be found below in figure 4.1 with details of the materials, structure and fabrication following thereafter.

4.1.2 Structure, composition and fabrication of hybrid AlGaInAs silicon-evanescent lasers

![Device structure of the hybrid AlGaInAs silicon-evanescent laser](taken from [8])

The silicon section of the device in which wave guiding takes place is fabricated from an undoped silicon-on-insulator substrate with a buried oxide thickness of 2 µm. A mask of SiO₂ is first deposited by plasma-enhanced chemical vapour deposition (PECVD), which is then patterned using electron-beam lithography to form a grating pattern. A Cl₂/Ar/HBr reactive ion etch (RIE) is then used to etch the regions exposed by the electron-beam, leaving a grating structure with a pitch of 238 nm such that the Bragg condition (equation 2.27, section 2.6.2) is met for λ=1600 nm emission wavelength in air. A further deposition of SiO₂ and RIE etch follows to define a ridge waveguide of height 0.76 µm and width 2.5 µm. When bonded this yields a single mode operation cavity with waveguide optical mode overlap of 59.2% and an active region optical confinement factor of 5.2% (for a multiple quantum well device) [9].

The single mode operation resulting from the quarter-wave shifted DFB grating structure of the waveguide has several benefits for a laser designed for use in an optoelectronic integration scenario. Firstly, for large bandwidth optical data transmission many signals of varying wavelength will be multiplexed to form a large data stream. For the data stream to be successfully demultiplexed the individual signals need to be distinguishable after transmission and thus the pulses which carry the signals need to be of narrow linewidth. Here the the DFB structure is ideal since linewidths of 3.6 MHz have been observed from this structure [9], allowing in excess of a hundred signals to be multiplexed together to achieve terabit per second data
Secondly, significant internal heating of electronic components is common, with CPUs operating at a temperature of up to 80°C. Thus for demultiplexing to be possible the individual signals need to not only exhibit a narrow linewidth but also good temperature sensitivity. The lasing wavelength in a standard Fabry-Perot cavity laser will follow the movement of the bandgap with temperature, changing by as much as 80 nm over this temperature range. The DFB cavity however forces the lasing to take place at the wavelength defined by the pitch of the grating, which will vary according to the thermal expansion coefficient of the silicon. This is far less than the change in the bandgap (since the two have a reciprocal relationship) and so the change in lasing line wavelength for the DFB laser would only be 8 nm at most in this range, meeting the requirements for multiplexing. A final advantage of this cavity type is that it does not require facets to act as mirrors in order to form a cavity, and as such is a lot more flexible in its placement and the properties of the surrounding materials that a Fabry-Perot laser.

The III-V section of the device consists of an AlGaInAs based MQW active region grown epitaxially on an InP substrate by MOCVD, full details of the structure are presented in table 1 below. AlGaInAs is the chosen active region material since firstly it can be engineered to emit around the low loss limit of silica fibres and waveguides of 1550 nm, the standard for communications. Secondly it has been shown previously [130] that AlGaInAs QW lasers exhibit a lesser temperature sensitivity than the conventional InGaAs/InP devices used in optical communications. As has been mentioned above internal heating is prominent in the type of electrical devices these lasers will be integrated within and as such a temperature insensitive threshold current will be essential in producing a system with a high overall electrical efficiency.

The strained superlattice is included just above the n-InP bonding layer in the in order to act as a blocking medium for threading dislocations formed at the bonding interface, which may propagate towards the active region, creating recombination centres and reducing efficiency. Since the extension of a threading dislocation tends to be energetically favourable once formed, they will tend to elongate over a long distance within a given material. However, an abrupt discontinuity in material structure, such as takes place repeatedly within a superlattice tends to inhibit threading dislocation propagation since the large mismatch induces a significant strain counteracting the strain field of the dislocations. Thus propagation into the superlattice is less energetically favourable than remaining in the original layer causing the dislocations to have a tendency to deviate away from the interface, protecting the regions beyond. It should however be noted that if the defect density is large enough the strain field can perturb the superlattice such that the strain in this region is reduced, decreasing the ability of the SSL to deviate the path of the threading dislocations [164].
<table>
<thead>
<tr>
<th>Material</th>
<th>Strain</th>
<th>Thickness (nm)</th>
<th>Doping</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>P cap</td>
<td>In$<em>{0.53}$Ga$</em>{0.047}$As</td>
<td>100</td>
<td>$10^{19}$ (P)</td>
<td>-</td>
</tr>
<tr>
<td>Cladding</td>
<td>InP</td>
<td>1500</td>
<td>$10^{18}$ (P)</td>
<td>3.167</td>
</tr>
<tr>
<td>SCH</td>
<td>Al$<em>{0.13}$Ga$</em>{0.34}$In$_{0.53}$As</td>
<td>125</td>
<td>$10^{17}$ (P)</td>
<td>3.562</td>
</tr>
<tr>
<td>QW (×8)</td>
<td>Al$<em>{0.09}$Ga$</em>{0.46}$In$_{0.45}$As</td>
<td>+0.85% (compressive)</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>Barrier (×9)</td>
<td>Al$<em>{0.05}$Ga$</em>{0.20}$In$_{0.65}$As</td>
<td>-0.55% (tensile)</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>SCH</td>
<td>Al$<em>{0.13}$Ga$</em>{0.34}$In$_{0.53}$As</td>
<td>125</td>
<td>$10^{17}$ (N)</td>
<td>3.562</td>
</tr>
<tr>
<td>N conduction</td>
<td>InP</td>
<td>110</td>
<td>$10^{18}$ (N)</td>
<td>3.167</td>
</tr>
<tr>
<td>Superlattice (A)</td>
<td>In$<em>{0.85}$Ga$</em>{0.15}$As$<em>{0.33}$P$</em>{0.67}$</td>
<td>-</td>
<td>7.5</td>
<td>$10^{18}$ (N)</td>
</tr>
<tr>
<td>Superlattice (B)</td>
<td>InP</td>
<td>7.5</td>
<td>$10^{18}$ (N)</td>
<td>3.167</td>
</tr>
<tr>
<td>Superlattice (A)</td>
<td>In$<em>{0.85}$Ga$</em>{0.15}$As$<em>{0.33}$P$</em>{0.67}$</td>
<td>-</td>
<td>7.5</td>
<td>$10^{18}$ (N)</td>
</tr>
<tr>
<td>Superlattice (B)</td>
<td>InP</td>
<td>7.5</td>
<td>$10^{18}$ (N)</td>
<td>3.167</td>
</tr>
<tr>
<td>Bonding layer</td>
<td>InP</td>
<td>10</td>
<td>$10^{18}$ (N)</td>
<td>3.167</td>
</tr>
<tr>
<td>SOI waveguide</td>
<td>Si</td>
<td>700</td>
<td>$10^{14}$ (P)</td>
<td>3.478</td>
</tr>
<tr>
<td>SOI insulator</td>
<td>SiO$_2$</td>
<td>1000</td>
<td>-</td>
<td>1.528</td>
</tr>
<tr>
<td>Substrate</td>
<td>Si</td>
<td>-</td>
<td>-</td>
<td>3.478</td>
</tr>
</tbody>
</table>

Table 4.1: The structure of the III-V mesa layers used to form the injection and active regions of the devices.

The III-V region is bonded to the SOI substrate using a low temperature oxygen plasma assisted flip chip wafer bonding technique [63]. As is to be expected surface quality is of the utmost importance for successful wafer bonding, requirements on surface flatness are fairly relaxed over the full extent of the wafer, with differences of the order of microns being acceptable since over long range the wafer can sufficiently deform to allow mating of the two surfaces. On a local scale however variation must be less than 1 nm otherwise voids may form. Clearly scratches, multistep discontinuities and surface contamination from large molecules such as organics and particles such as dust may all therefore cause voids if present on the surface.

To prepare the surfaces the wafers first each undergo a chemical surface cleaning treatment each, HF and NH$_4$OH in the case of the Si and InP surfaces respectively [8]. Wafer bonding of individual chips after chemical treatment alone is possible however bonding of entire wafers was found to lead to shattering and debonding [72]. The issue here was that high anneal temperatures were required (630 °C) and since in addition to their large lattice mismatch Si and InP also differ in thermal expansion coefficient by 2.15 μm °C$^{-1}$ leading to a build up of strain greater than that which the lattice can accommodate.

To combat this a further step of oxygen plasma surface treatment for both wafers simultaneously is carried out. Oxygen is known to be of particular use in the removal of water and organics, and since in previous bonding trials voids and water bubbles have caused major defects in the bonded system the removal of these species is essential. The exact mechanism by which the oxygen plasma assists in bonding is a point of contention but it has been proposed that the observed oxide buildup on the Si surface produces highly strained Si-O-Si bonds which are therefore particularly reactive. Since both wafers are exposed to the plasma simultaneously within the RIE chamber they can be bonded in situ without exposure to the ambient air maintaining surface cleanliness. Once in contact the wafers are placed under 1.5 MPa of uniaxial pressure for 12 hours at an annealing temperature of 300 °C. It was found that although significant strain is still generated as a result of the thermal expansion coefficient mismatch the post bonding stress tests led to fracturing of the InP substrate rather than the bonding region, indicating that the bond is at least as strong as the substrate.
Once bonded chemical wet etching is used to define the mesa structure depicted in figure 4.1 above, and metallic contacts deposited on the exposed n-type InP either side of the mesa. Next proton bombardment of the mesa is used to and define a 4 $\mu$m wide current channel, which ensures that a uniform current density passes through the active region above the waveguide, where the optical confinement factor of the quantum wells is maximised. The mode is therefore gain guided laterally within the III-V region. Finally p-type contacts are deposited and the wafer is diced into bars of completed devices.

![Hybrid silicon platform](image)

Figure 4.2: Cross-sectional representation of the Hybrid structure along side a TEM image of the active region. The SSL layer is observed to block the propagation of defects into the active region (image modified from original in [133]).

Photoluminescence studies were carried out at UCSB on post bonding wafers [133], with and without the addition of a SSL layer using a 980 nm, 0.5 mW pump laser at room temperature. These yielded the result that with the addition of the SSL layer, the photoluminescence intensity detected from the wafers increased by a factor of 7 as shown in figure 4.3 below. Since TEM studies indicated the effective blocking of the propagation of defects produced during the wafer bonding process (figure 4.2) it is indicated that a reduction in defect density within the active region has led to an increase in radiative efficiency.
4.2 Distributed feedback AlGaInAs-silicon evanescent laser characterisation

The inclusion of a strained superlattice (SSL) in the proximity of a region for which material quality is paramount has been previously shown to prevent the propagation of defects such as threading dislocations provided the defect density is not excessive [56]. The research group responsible for the AlGaInAs-silicon evanescent laser have demonstrated the blocking of defect propagation resulting from wafer fusion (See figure 4.2) as well as a resulting improvement in photoluminescence intensity (figure 4.3).

Two structures are investigated here, differing only in that one contains a strained superlattice defect blocking layer and the second does not. These two structures are compared throughout the following investigation so as to ascertain the physical effects of introducing such a layer. The structure of the III-V active region is presented in table 1, with the complete device cross-section depicted in figure 3.1. For the purposes of threshold current density calculations the dimensions of the pumped area of the active region are; length: 280 $\mu$m and width: 4 $\mu$m, giving an area of $11.2 \times 10^{-6}$ cm$^2$.

4.2.1 Electroluminescence

Initially the threshold current densities of all available devices were determined through electroluminescence using the probe station system described in sections 3.1 and 3.2.1. The histogram depicted in figure 4.4 shows a clear difference in the threshold current density distribution between those devices which contain a SSL and those which do not. Counter to expectations that the SSL would reduce or at worst have no effect on
threshold current density the inclusion of such a layer is actually demonstrated to increase the mean threshold current density from $3.39 \text{ kAcm}^{-2}$ to $4.91 \text{ kAcm}^{-2}$ as well as broadening the overall distribution. It does however appear that the SSL layer leads to $\sim$ twice the yield of functional devices. These values represent threshold current densities per quantum well of $423 \text{ Acm}^{-2}$ and $614 \text{ Acm}^{-2}$. Regardless of the structure the observed current densities are far greater than those observed from the industry standard InGaAs/InP lasers where current densities of the order of $100 \text{ Acm}^{-2}$ per quantum well can be expected [113]. Thus significant efficiency limiting processes must be taking place in these devices regardless of the presence of the SSL in order to achieve such large threshold current densities.

Since the threshold current has a large standard deviation of 20.7 mA and 6.8 mA for the SSL and non-SSL containing devices respectively, it is reasonable to assume that growth quality and/or processing quality across the wafer is inhomogeneous and that this would be to a similar extent in both samples. Should this be the case and the SSL prevents defects reaching the active region it is to be expected that only the best of the non-SSL containing devices would be operational. Conversely with a lower defect density in the active region on average, inherently lower quality SSL containing devices could be operational, extending the threshold current density distribution to higher values. Again, with twice as many operational devices it is likely that the SSL containing subset is a more representative view of the overall device quality distribution.

Figure 4.4: The binned distribution of device threshold currents for both SSL and non-SSL containing structures.

From the sample of operational lasers a representative device for SSL and non-SSL containing structures was selected for further investigation. Below in figures 4.5 and 4.6 the temperature dependent light intensity vs current plots (LIs) for the SSL and non-SSL containing devices respectively are presented. Electroluminescence was carried out under pulsed operation with a frequency of 10 kHz and pulse width of 500 ns, giving a duty
cycle of 0.5% in order to minimise Ohmic heating to a negligible level as this has been found to cause rapid thermal rollover from CW characterisation within the Bowers group at UCSB [8].

Figure 4.5: The L-I curves for a representative laser featuring a SSL defect blocking layer.

Figure 4.6: The L-I curves for a representative laser which does not feature a SSL defect blocking layer.

The SSL containing structure (figure 4.5) was found to lase up to a maximum temperature of 50 °C with a current density of 10.6 kAcm$^{-2}$, in comparison to the non-SSL containing structure (figure 4.6), which just about lased up to 60 °C with a current density of 9.7 kAcm$^{-2}$. The greater maximum operating temperature and lower threshold current densities at each temperature again suggest that the inclusion of the SSL layer is
not of benefit in terms of threshold current density and in fact actually increases the value. Also in contrast
to the findings that photoluminescence intensity is increased by a factor of 7 by the inclusion of a SSL layer
[133], the electroluminescence at 10 °C is improved by 20%, eventually falling behind that of the non-SSL
containing structure as temperature increases.

Peak external quantum efficiency or slope efficiency at ambient temperature is found to be 21% and 14% in
the cases of the SSL and non-SSL containing structures respectively. Both of these values are particularly low
in comparison to commercial GaAs based devices where efficiencies approaching unity are readily achievable.
Thus, independent of the effects of the SSL, significant reduction in efficiency limiting processes is required
to achieve a commercially viable laser.

Comparing the two structures it is evident that at all temperatures the presence of the SSL leads to an
increased slope efficiency. Since the expected effect of the SSL is to reduce the defect density but have little
other effect it is counter intuitive that the slope efficiency would improve since recombination mechanisms
should have no effect on the slope efficiency aside from the case of a failure of the carrier density to pin. The
other alternative causes would be either an increase in internal efficiency, as given by equation 2.36 caused
by an improvement in the efficiency with which carriers are injected into the active region or a decrease in
optical losses (see equation 2.34). From table 4.1 it can be seen that the n-type conduction layer is found
immediately above the SSL, which is in turn only 10 nm from the bonding interface, or in the case of the
non-SSL devices just above the 10 nm bonding layer. Threading dislocations will have to pass through this
region in order to reach the active region and so it is reasonable to expect a larger density of defects in this
current injection layer in the non-SSL devices. This would cause these to have a lower injection efficiency
than the SSL containing devices, as carriers injected from the n-contacts could interact with these states and
become trapped, lowering the internal efficiency and fitting with the observations from figures 4.5 and 4.6.
There is not much information from these plots alone to determine if a drop in optical losses is the cause
however it is difficult to imagine how the addition of the SSL would have much effect on this type of loss.

Whether this reduction in defect density is beneficial solely through the reduction of trap density in the
n-type injection layer or if additional loss mechanisms are also suppressed is unclear from this information
alone however. That the temperature sensitivity of the slope efficiency is lower with the addition of the
SSL is indicative of a subsequent reduction in a loss mechanism with a greater temperature sensitivity than
radiative recombination. Thus this must be either Auger recombination or carrier leakage. Optical losses
could also be responsible for a falling slope efficiency with temperature since holes will have an increased
thermal energy and therefore be able to access deeper states in the valence band and increasing the cross
section for IVBA.

The SSL containing device is also superior in slope efficiency as a function of current. Up to the limit
of 150 mA the LI plots for the SSL containing structure are fairly linear above threshold, maintaining their
peak differential quantum efficiency. In contrast the external quantum efficiency of the non-SSL containing
structure rapidly drops off with increasing current. Since thermal rollover is unlikely due to the low duty cycle,
it is implied that one or more non-radiative recombination mechanisms is taking place that has a stronger
dependence on the carrier density than radiative recombination ($\propto n^2$). The process could be dependent on applied bias voltage, since this is also increasing, but the fact that slope efficiency drops off at an even greater rate for higher temperatures backs up the hypothesis that carrier density is the dominant dependency. It should be noted here that a change in recombination mechanism fractions leading to a reduction in slope efficiency above threshold necessarily requires the carrier density to fail to pin at threshold, otherwise the fractions will be fixed (see equation 2.21). This is observed in section 4.2.2 and with its impact explained in more detail alongside the relevant data.

4.2.2 Spontaneous emission

The pure spontaneous emission (SE) from the devices was collected using a multimode optical fibre placed in contact with the substrate, as shown in figure 3.2 of section 3.2.2. Since the substrate is made from the same material as the waveguide (Si) it is therefore transparent over the range of wavelengths emitted through spontaneous emission, as is the SiO$_2$ insulating layer, which can be seen to be highly transmissive and feature free in the region of the electromagnetic spectrum of relevance from figure 4.7 below. Thus the SE can be collected from the underside without the need for milling or other alterations to be made to the devices.

![Optical transmission spectrum SiO$_2$](image)

Figure 4.7: Optical transmission spectrum SiO$_2$ [96]

Figure 4.8 shows the spontaneous emission intensity as a function of current, the threshold currents are marked with circles. The first point of note here is that whilst the facet emission power is at most 20% larger and eventually less with the addition of a SSL, in the case of the SE power it is a factor of 5 greater. Since the photoluminescence is closely related to the spontaneous emission it is logical that a similar increase is found in both with the addition of the SSL. From section 2.2.2, equation 2.9 an increase in spontaneous emission at a given current implies that the $Bn^2$ term is proportionately larger than the competing non-
radiative components. It would be expected that an increase in spontaneous emission power would lead to a proportionate increase in the stimulated emission power, since the stimulated emission rate (see equation 2.10) is proportional to the photon density within the cavity. However, as was observed from figures 4.5 and 4.6 this is not the case. These two results appear to contradict each other, and so, there must be an additional impact of the inclusion of the SSL beyond the suppression of non-radiative recombination mechanisms. A second immediately obvious result is that carrier density pinning does not take place in the case of either structure under investigation. Initially, the lack of pinning was attributed to scattered stimulated emission, which would be expected to continue to increase with current.

![Figure 4.8: The spontaneous emission collected from the base of the non-SSL containing (left) and SSL containing (right) structures as a function of current. The threshold currents are marked by circles.](image)

Linear interpolation was used to produce an approximation of the spectra excluding this scattered light as a function of current. Referring to figure 4.14 in section 4.2.5 scattered stimulated emission is clearly passing through the substrate and being detected in addition to the spontaneous emission. The peaks however are very sharp as a result of the DFB cavity and so if the data at the lasing peak wavelength is removed, the interpolation across the gap representing the spontaneous emission alone is fairly simple and requires little assumption. Doing so for several currents at each temperature gives a set of spectra that have been “corrected”, having the scattered stimulated emission removed. Integrating these spectra gives a value at each temperature and current of integrated spontaneous emission, all be it with a different collection efficiency to the data plotted in figure 4.8 due to having to use an OSA rather than an optical power meter to detect the light. Normalising the low injection (where the stimulated emission rate is negligible and no scattered signal is discernible (e.g. as shown in figure 4.9)) integrated spectra values to the integrated spontaneous emission powers in figure 4.8 allowed the substrate emission luminosity to be corrected in order to show only contributions from pure spontaneous emission. The spontaneous emission data has been replotted in figure 4.9, along with the corrected data points at 10mA intervals. Interestingly it can be seen that after correction the SSL and non-SSL containing structures exhibit some pinning behaviour below 30 °C and 40
as the current continues to increase the corrected data asymptotes towards the original data sets, indicating that the pinning is rapidly degraded as a function of current. Therefore the efficiency limiting mechanism(s) which cause the devices to stop pinning is both temperature and current dependent. Comparing the SSL and non-SSL containing structures it can be seen that the addition of an SSL is detrimental to pinning, lowering both the temperature and current up to which pinning takes place.

Figure 4.9: The spontaneous emission collected from the base of the non-SSL containing (left) and SSL containing (right) structures as a function of current after correction by removal of scattered stimulated emission. The threshold currents are marked with grey circles.

If the spontaneous emission luminosity is not pinning above threshold the implication is that gain must also be continuing to increase. From equation 2.32 in section 2.6.3 it can be seen that the only parameter on which gain is dependent, which will be increasing consistently with current, is the carrier density. In an ideal device, above the lasing threshold the stimulated emission process would be expected to exhibit a recombination rate equal to the carrier injection rate, thus maintaining a constant carrier density. The lack of spontaneous emission pinning here implies that this is not happening in these devices.

From sections 2.2 and 2.3 it can be seen that the possible recombination mechanisms that may take place in a semiconductor laser have a varying dependency on carrier density; SRH recombination is \( \propto n \), radiative recombination is \( \propto n^2 \), Auger recombination is \( \propto n^3 \) and carrier leakage is \( \propto \exp\left(-\frac{E_B}{k_B T}\right) \). Thus, if the carrier density continues to rise, the proportion of carriers (injected current) consumed through Auger recombination, and potentially carrier leakage will also rise. The result is a decreasing external efficiency as a function of current.

The implications of this are twofold. Firstly the peak output power of the device will be limited, since diminishing returns of lasing light from the injected current will lead to an asymptotic LI curve and greater than necessary electrical power consumption. Secondly recombination via Auger mechanisms leads to the dissipation of energy through phonons rather than emitted photons. The phonon energy is added to the
lattice and observed as a heating effect on the system, in addition, carrier leakage increases ohmic heating. Again, from section 2.3 we see that both carrier leakage and particularly Auger recombination have a strong dependence on temperature. The resulting increase in temperature therefore amplifies the effects of these non-radiative process, further reducing efficiency, and leading to additional heating effects. This positive feedback between temperature and efficiency may lead to thermal roll-over, where the slope efficiency becomes negative, limiting the maximum operating current and optical output power.

A final point of note here is the disproportionately large reduction in slope efficiency between 40 and 50 °C and 30 and 40 °C in the cases of the non-SSL and SSL containing structures respectively. This sudden drop in slope efficiency is indicative of the presence of a temperature dependent, efficiency limiting process in both structures but whose activation temperature is reduced by the presence of the SSL. This difference is mirrored in the maximum temperature up to which pinning takes place, indicating that this mechanism becomes dominant fairly abruptly and is common to both structures, regardless of the SSL but is potentially exacerbated by its presence.

4.2.3 Characteristic temperature

Further analysis of the LIs from section 4.2.1 allows the characteristic temperatures, $T_0$ and $T_1$ to be derived from the threshold currents and external efficiencies according to the formulae found in section 2.10. Firstly, the characteristic temperature, $T_0$ of the non-SSL device, shown in figure 4.10 has an average value of $54\pm20$ K, considerably less than the temperature at which the experiment took place. By comparison to the characteristic temperatures for lasers dominated by each of the recombination mechanisms, it can be seen that this device is certainly not radiatively dominated or defect recombination dominated, since values of 300K and 200K would be expected in these cases respectively. An Auger recombination dominated device however would be expected to exhibit a value $\sim50$K, in close agreement with the experimental findings. For pure carrier leakage the characteristic temperature will be again dependent on $E_B$, which is dependent on both temperature and current density, as it is a comparison between band filling and the barrier height. Since the band filling is somewhat difficult to calculate and transient there is no quantified pure carrier leakage characteristic temperature. There is therefore no value available for a direct comparison to the $T_0$ value. It should be noted that there is a slight change in gradient at $\sim35$ °C in both $T_0$ and $T_1$ for both SSL and non-SSL containing devices. If $T_0$ is evaluated above and below this "kink" in the plot separately values of $105\pm15$ K below $35$ °C and $34\pm15$ K above $35$ °C are found. The sharp drop in $T_0$ is an indication of the onset of a temperature sensitive loss mechanism. Neither 105 K nor 34 K align with Auger recombination as the average value did, or with defect and radiative recombination either. Therefore since carrier leakage can take any value it would be possible that it could take a low enough value to pull the $T_0$ value below 50 K. The larger value of $105\pm15$ K would have to result from a combination of mechanisms that may well include both significant Auger recombination and carrier leakage to have such a low value.

The value of $T_1$ is also low, with an average across the temperature range of only $24\pm26$ K, with separate values below and above the "kink" of $53\pm21$ K and $13\pm21$ K respectively. The particularly low value above
the discontinuity suggests a strongly temperature dependent loss mechanism, such as a Auger recombination or carrier leakage dominates these devices, fitting with the $T_0$ findings.

Figure 4.10: The $T_0$ and $T_1$ variables plotted as a function of temperature for the non-SSL containing structure.

The structure containing a SSL defect blocking layer exhibits an average characteristic temperature, $T_0$, slightly less than the non-SSL structure of $48\pm20$ K, compared to $54\pm20$ K. Since the difference is so small it cannot be concluded with a reasonable degree of confidence that the introduction of the SSL had a significant effect. Thus it is apparent that the mechanism causing the particularly temperature sensitive threshold current is common to both structures and as such is not suppressed by the SSL defect blocking layer. If the separate gradients above and below $35^\circ$C are compared it is found that below $35^\circ$C the SSL value is actually lower, at $64\pm15$ K, compared to $105\pm15$ K, with both structures exhibiting similar values of $38\pm15$ K and $34\pm15$ K above. Since the only values that differ enough to be considered statistically significant are the lower temperature subsection this could potentially indicate that the non-SSL containing device is actually superior below $35^\circ$C, but once the additional hypothesised temperature activated mechanism is activated the two exhibit similar performance, suggesting that this mechanism is slightly suppressed by the SSL. This hypothesis of course will require data from the analysis that follows to be verified.

The average $T_1$ value is also similarly low at $31\pm26$ K, which would indicate that the mechanism limiting $T_1$ is also common to both structures. However in the two separate regions either side of the observed discontinuity, it is observed that the low temperature $T_1$ is significantly greater, with a value of $118\pm21$ K, compared to $53\pm21$ K whilst the higher temperature value is approximately the same as in the non-SSL case. There are two implications from these findings. Firstly that below the activation temperature of the proposed thermally activated efficiency limiting process the SSL defect blocking layer has a significant impact
on external quantum efficiency. This would be consistent with the expected result that the SSL would block the propagation of defects into the active region and thus improve efficiency by reducing the loss of injected carriers through SRH recombination, since this would be expected to be a more dominant process at low temperatures. Secondly, because the two structures have similar $T_1$ values above the activation temperature the thermally activated loss mechanism must be highly temperature dependent and become a dominant loss mechanism in comparison to the relatively temperature insensitive SRH recombination mechanism. The results of the $T_1$ comparison then completely contradicts the results of the $T_0$ comparison, again further investigation is required to determine the effect of the SSL on these parameters. It is however clear that both devices are limited by highly temperature sensitive loss mechanisms with both Auger recombination and carrier leakage looking to be significant.

Figure 4.11: The $T_0$ and $T_1$ variables plotted as a function of temperature for the SSL containing structure.

4.2.4 Z analysis

Z analysis was carried out as a function of temperature according to the method outlined in section 2.9. The results are depicted below in figure 4.12, where the blue data represents the non-SSL containing structure and the black data the SSL containing structure. $Z$ was calculated at both threshold (square) and at half the threshold (triangular) in order to indicate the dependence on current of the dominant recombination mechanisms. Since the Z analysis gives an indication of the average carrier density dependence of the net recombination taking place as a result of the sum of all mechanisms present it should give a further indication of which mechanisms are dominant in these devices.
Figure 4.12: Z values for both structures as a function of temperature calculated at both threshold injection current and half this value. Black and blue data represent SSL and non-SSL containing structures respectively; whilst square and triangular data indicate threshold and half threshold injection currents respectively.

As a function of temperature, the low current Z value rises slowly in the case of the SSL containing structure, from 2.18 at 10°C to 2.28 at 30°C at a rate of 0.005 K⁻¹. At 30°C, where a discontinuity has been observed in the previous sections the rate of increase of Z increases to 0.009 K⁻¹, rising to 2.46 at 50°C. These values are close to the ideal value of 2, representing pure radiative recombination, suggesting that at low injection and temperature the SSL containing devices are dominated by radiative recombination and are therefore fairly efficient. This, however does not match with the threshold current densities or slope efficiencies of the devices and so seems fairly unlikely, with a combination of SRH and Auger recombination (Z=1 and Z=3) happening to give a value around 2. The discontinuity, as observed in the studies of spontaneous emission and LIs although small can still be observed here, increasing the Z value and implying the rapid onset of a thermally activated loss mechanism with a Z value greater than 2, i.e. Auger recombination or potentially carrier leakage.

In the case of the low injection non-SSL containing structure the Z value at 10°C is found to be 1.8, rising to 2.13 at 40°C, giving a rate of 0.011 K⁻¹. Again, in this case more prominently the discontinuity can be seen, with the rate of increase of Z with temperature rising to 0.2 K⁻¹ giving a value of 2.51 at 60°C. The Z values of less than 2 imply the presence of defects, since defect recombination is the only mechanism with a characteristic Z value of less than 2. The implication here is that the SSL is having the predicted effect of suppressing the propagation of defects into the active region and therefore minimising SRH recombination. Z as a function of temperature tends towards the values observed from the SSL containing structure, as is to be
expected due to the more thermally sensitive recombination mechanisms taking over from the temperature insensitive defect recombination contribution.

Using the mathematical treatment presented in section 2.9.2 of chapter 2 the Z values can be fitted by varying fractions of SRH, radiative and Auger recombination (carrier leakage is assumed negligible in this treatment, valid at low injection). The possible combinations that give Z values of 1.8 and 2.18 are plotted below with solid and dotted lines for the non-SSL and SSL containing structures respectively. It can be seen that these two values can be achieved for radiative fractions between as much as 0.8 and as little as zero, for varying Auger and defect fractions, this indicates that the absolute value of Z alone cannot always be taken at face value in comparison to the characteristic values as there are many ways to achieve each value. As will be detailed in the following paragraph the defect fraction at threshold is estimated to be 29±2% and 40±2% at 10 °C for the SSL and non-SSL containing structures respectively. Adding these limits to figure 4.12 narrows the range which the radiative and Auger fractions may take, as this data is for a low injection regime it is likely that the defect fraction is actually greater than at threshold due to the linear dependence on carrier density of this mechanism compared to quadratic or cubic for radiative and Auger recombination respectively, but serves as a reasonable approximate lower limit for the SRH fraction. With these in place the maximum radiative fractions are now only 24% and 40%, leaving 47% and 20% Auger recombination for the SSL and non-SSL containing devices. These values of course represent the limit of the lowest possible defect fraction and assume no carrier leakage, making them no more than an indication of the approximate levels of each recombination mechanism that are feasible.
Figure 4.13: Possible combinations of SRH, radiative and Auger recombination that can be used to fit the $Z$ values close to 2 seen at low injection, suggesting that these values do not represent a radiatively dominated device.

When considering the threshold $Z$ values (squares) it can be seen that at low temperature the $Z$ value of the SSL containing structure is 3.47 with a slope of $0.01 \, K^{-1}$, compared to 3.23 with a slope of $0.006 \, K^{-1}$ in the case of the non-SSL structure. So as in the case of low injection the $Z$ value is larger in the case of the SSL containing structure at low temperature, this could be a result of a larger defect recombination contribution in the non-SSL containing structure reducing the $Z$ value. By extrapolating from the low current regime, where a $Z \approx 1$ region can be found (i.e. almost completely dominated by defect recombination) to threshold gives an estimation of the defect recombination percentage. The full details of this approximation are presented in section 2.9.1. The extrapolation gives the percentage of current density accounted for by defect recombination and is summarised for each structure and temperature in table 2 below. It was found that defect related recombination accounted for $40\pm2\%$ of the threshold current at $10^0C$ in the case of the non-SSL containing structure, whereas in the case of the SSL containing structure this is only $29\pm2\%$. Thus a $\sim7\%$ increase in $Z$ value with the addition of a SSL layer corresponding to a $11\%$ decrease in defect current fits well with the expectation of SSL layers hindering the progression of threading dislocations into the active region. It is also quite possible that carrier leakage is taking place to a greater degree in the SSL containing structure as a result of the elevated threshold carrier density causing greater band filling and therefore a greater fraction of carriers to have enough energy to escape the confinement of the well. The reality is therefore likely a combination of the two effects.

As temperature is increased it can be seen that the $Z$ value of the SSL containing structure increases to
4.16 at 50°C, with a slope of 0.024 K⁻¹. A strong increase with temperature giving values well beyond 3 indicates that in addition to Auger recombination carrier leakage must be present in order to give such a large and rapidly increasing value of Z. In contrast the non-SSL containing structure rises to 4.17 by 60°C with a slope of 0.037 K⁻¹. Thus initially it seems that since the Z value is consistently lower at each temperature the degree of carrier leakage is lower in the non-SSL containing structure. However it should be taken into account that the threshold currents at each temperature are lower for the non-SSL containing structure and so the band filling is lower, leading to less leakage. The rate of increase of Z is 0.004 K⁻¹ lower at 10°C but by 60°C it is 0.013 K⁻¹ greater. Thus, as a function of temperature the “acceleration” of growth of the Z value with temperature for the non-SSL containing structure is greater, implying that the temperature dependent loss mechanism responsible is stronger at a given current.

Comparing the low to high injection regimes (triangles to squares) it is evident that the Z value is highly dependent on the injection current density. Increasing the injection current to threshold causes the carrier density to rise also, and therefore modify the fractional contribution of each recombination mechanism due to their varying dependence on carrier density. Thus it is to be expected that the Z values tend towards the characteristic value of the higher carrier density dependence mechanisms, i.e. Auger recombination and carrier leakage. Carrier leakage in particular, since it is dependent on the band filling level and so would be expected to increase rapidly with injection current, as is observed.

Since both exhibit Z values greater than 3 and continue to increase strongly, a recombination mechanism with a greater than cubic carrier density must be present. Since SRH, radiative and Auger recombination have linear, quadratic and cubic carrier density dependencies respectively the cause of the strongly temperature dependent and large Z value must be, at least partially, caused by carrier leakage. Thus from the Z analysis there is a strong indication that both defect recombination and carrier leakage are taking place to a significant degree, since values of Z < 2 and > 3 have been observed at low and high injection currents respectively. The presence of Auger recombination is possible and given the strongly temperature dependent threshold current and external quantum efficiency is likely to be taking place to some degree, but since it is represented by an intermediate value of Z, it is not possible to determine what proportion of total recombination it accounts for from this Z analysis alone. In section 5.2.1 of chapter 5 the Z analysis is used in conjunction with simulation to determine the relative proportions of each mechanism taking place in these devices. This cannot be carried out in this case mainly due to the fact that the carrier leakage is particularly difficult to simulate since the band filling and precise band alignments are required. In addition, the radiative component is similarly difficult to determine as pinning does not take place and even at low injection it cannot be said that the devices are completely radiatively dominated.

The results of the threshold defect fraction approximation by extrapolation are summarised in table 4.2 below.
Table 4.2: A summary of the threshold currents and fractional defect current contributions as a function of temperature for both structures.

<table>
<thead>
<tr>
<th>Temperature °C</th>
<th>$I_{th} (mA)$ [No SSL]</th>
<th>$I_d$ [No SSL]</th>
<th>$I_d / I_{th}$ [No SSL]</th>
<th>$I_{th} (mA)$ [SSL]</th>
<th>$I_d$ [SSL]</th>
<th>$I_d / I_{th}$ [SSL]</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>42</td>
<td>17</td>
<td>40 ± 2%</td>
<td>52</td>
<td>15</td>
<td>29 ± 2%</td>
</tr>
<tr>
<td>20</td>
<td>45</td>
<td>18</td>
<td>39 ± 2%</td>
<td>59</td>
<td>16</td>
<td>28 ± 2%</td>
</tr>
<tr>
<td>30</td>
<td>51</td>
<td>19</td>
<td>37 ± 2%</td>
<td>71</td>
<td>19</td>
<td>28 ± 2%</td>
</tr>
<tr>
<td>40</td>
<td>61</td>
<td>21</td>
<td>35 ± 2%</td>
<td>89</td>
<td>20</td>
<td>22 ± 2%</td>
</tr>
<tr>
<td>50</td>
<td>75</td>
<td>23</td>
<td>30 ± 2%</td>
<td>119</td>
<td>24</td>
<td>20 ± 2%</td>
</tr>
<tr>
<td>60</td>
<td>109</td>
<td>26</td>
<td>24 ± 2%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

It should be noted at this point that Z analysis assumes that defects are uniformly distributed throughout the active region. The wafer bonding method by which the hybrid lasers are fabricated leads to an initially high crystalline quality active region being bonded to a Silicon substrate at an interface from which a large density of threading dislocations will extend [53]. Therefore if these reach the active region the geometry of the defect distribution may be approximately planar, counter to our assumptions. Whether this would have a significant effect or not would depend on exactly how far the threading dislocations extended into the active region, if they pass through the bulk of this region then the distribution can again be thought of as uniform. At this stage it is thought that the effect should not be significant to such a degree that the Z values are skewed significantly enough to be unreliable. This is because these are used mainly as an indication of the dominant recombination mechanism in this analysis and only one fractional threshold component (SRH recombination) is calculated explicitly. This value is also only used as a comparison between two devices which exhibit the same geometry in their structure, in terms of the way in which defects will propagate in a planar manner towards the active region.

Z analysis has indicated that both structures have a significant temperature dependent loss mechanism(s) present, which is strongly indicated to be carrier leakage by the Z value of >3. Auger recombination may be present but to what degree is difficult to quantify from these results alone since it is an intermediate value of Z and so cannot easily be separated from the competing recombination mechanisms. That carrier leakage (and potentially Auger recombination) is a dominant loss mechanism is consistent with the previous results of low characteristic temperatures, $T_0$ and $T_1$, a non-pinning of the carrier density at threshold and a decreasing external quantum efficiency with current. The low injection regime Z values suggest that the defect recombination contribution to the total threshold current is ~30% and ~40% in the case of the SSL and non-SSL containing structures respectively. It seems that although the SSL layer is reducing the contribution from SRH recombination, it remains significant enough in both structures to have a limiting effect on efficiency.

Comparison of the Z values of the two structures indicates that the addition of the SSL does reduce carrier leakage, although this is somewhat masked by the fact that the threshold current density increases. Again this is consistent with the previous results of a larger “pre discontinuity” $T_1$ value and a larger slope efficiency in the case of the SSL containing structure. It has been previously demonstrated [110] that a region of sufficiently high defect density can act as a carrier leakage path, this would be consistent with the effect of the SSL on the active region material quality as other than through defect density reduction the SSL should
not affect the carrier confinement of the QWs.

4.2.5 Spectral analysis

Spectra were produced as a function of temperature and injection current in tandem with the electroluminescence investigation detailed in the previous sections of this chapter. Figure 4.14 shows those produced below threshold at an injection current of 20 mA under the same pulsed conditions used to collect the spontaneous emission and LI plots from sections 4.2.2 and 4.2.1. Spectra were produced in this manner to maintain consistency when comparing spectral data to other types. However, the spectra of both the SSL and non-SSL containing devices were weak giving a low signal to noise ratio (SNR) and requiring a 40 pt smoothing to be used to allow the peaks to be clear. A further result of the weak signal is that a large current was required to produce a spectrum at higher temperatures. For a fair comparison an equal current is used at each temperature and so some scattered amplified spontaneous emission is collected in the lower temperature measurements, where the threshold currents are lower.

The presence of amplified stimulated emission indicates that the device is approaching lasing threshold and that stimulated emission is becoming a prominent recombination mechanism. As such the spectrum in this vicinity could potentially be perturbed and so not indicative of sub-threshold spectral behaviour. The complex structure of these devices can reasonably be expected to lead to similarly complex spectra and as a result the spectral results were repeated under continuous wave (CW) operation, increasing the light emission intensity. However, as result, ohmic heating will also have increased proportionately, leading to the previous assumption that this effect is negligible to no longer valid. The spectra produced under CW injection therefore, will be perturbed in comparison to those produced under pulsed injection, particularly at the largest injection currents.

From figure 4.14 several prominent features are observed. There appear to be three peaks centered on ∼1370 nm, ∼1490 nm and ∼1600 nm. Of these the peak around 1490 nm is the largest and as a result was initially assumed to indicate the approximate position of the peak of the gain spectrum (see section 2.5.2) of the active region. However, contrary to this observation, a narrow peak (broadened by the smoothing) attributed to scattered amplified stimulated emission is found to peak around 1600 nm, suggesting that lasing will take place at this wavelength and therefore the peak gain should be around this wavelength.
Figure 4.14: Spectra of the emission from the substrate of both structures as a function of temperature when pumped using a pulsed injection current of 20 mA (below threshold). Left: SSL containing. Right: non-SSL containing.

Repeating the temperature dependent spectral measurements below threshold at 10 mA CW injection yields the plots shown in figure 4.15 with the SSL and non-SSL containing structures on the left and right respectively. With a larger duty cycle and emission intensity the peaks can be seen more clearly, particularly in the case of the non-SSL containing structure, which had an almost undetectable spectrum under pulsed operation. The scattered stimulated emission detected previously is also not observed since a lower injection current can be applied with the significantly larger duty cycle.

In the case of the SSL containing structure the peak at ∼1490 nm decreases in intensity relative to the second peak at ∼1600 nm as well as strongly red shifting from ∼1540 nm to ∼1590 nm between 10 and 60 °C. The temperature dependent behaviour of the spectra in the case of the non-SSL containing structure follows a similar pattern but with the first peak blue shifted by 10 nm, at around 1480 nm. These spectra imply two separate emission sources differing by up to 0.06 eV, with the lower energy (longer wavelength) emission source being highly temperature sensitive with $d\lambda/dT = 1\text{nmK}^{-1}$ compared to only $0.2 \text{nmK}^{-1}$ for the higher energy (shorter wavelength) peak. The third peak, centering on ∼1370 nm and ∼1350 nm in the case of the SSL and non-SSL containing structures respectively remains weak and difficult to accurately determine the position of.
Figure 4.15: Spectra of the emission from the substrate of the SSL containing (left) and non-SSL containing (right) structures as a function of temperature when pumped using a CW injection current of 10 mA (below threshold).

Increasing the injection current to 120 mA, under CW operation, produced the spectra shown in figure 4.16 with the SSL and non-SSL containing structures on the left and right respectively. The central peak moved by only a few nm as a result of increasing the current above threshold, comparable to the uncertainty in the measurement and so is not considered a significant effect in this analysis. The second peak, however redshifts by up to \( \sim 60 \) and \( \sim 45 \) nm as well reducing \( \frac{d\lambda}{dT} \) from 1 \( \text{nmK}^{-1} \) to 0.34 and 0.4 \( \text{nmK}^{-1} \) in the case of the SSL and non-SSL containing structures respectively. Additionally, scattered stimulated emission from the DFB waveguide is detected producing a series narrow and intense peaks in the vicinity of the second peak at \( \sim 1600 \text{ nm} \) with FWHM of 1nm and \( \frac{d\lambda}{dT} \) of 0.08 \( \text{nmK}^{-1} \). The third peak, which below threshold and in pulsed operation was poorly defined is now clearly resolved, centering on \( \sim 1360 \) and \( \sim 1350 \) nm in the case of the SSL and non-SSL containing structures respectively with \( \frac{d\lambda}{dT} \) = 0.1 \( \text{nmK}^{-1} \) and can be seen to now have a comparable intensity relative to the second peak. Another interesting point of note is that the relative height of the central peak compared to the satellite peaks is \( \sim 2 \) in the case of the SSL containing structure but only \( \sim 1.4 \) in the case of non-SSL containing structure. This indicates that as well as the total integrated intensity being improved by the presence of the SSL, as was found in section 4.2.2, the distribution of emission across the three hypothesised distinct emission sources is dependent on the presence of the SSL also.
The positions of each peak as a function of temperature are plotted in figure 4.17 for both the SSL (left) and non-SSL (right) containing structures, including a comparison of pulsed to CW injection both above and below threshold. In these plots black/blue symbols signify an injection current below/above threshold, solid/hollow symbols signify pulsed/CW injection and square/triangle/circle/star symbols signify the central/second/lasing/third peak position.

Firstly comparing the solid to hollow symbols allows the effect of increasing the duty cycle from 0.5% (pulsed) to 100% (CW) to be determined. Across both structures the peak emission wavelengths asymptote to single values across this temperature range making adjusting for this perturbation simpler when comparing pulsed to CW spectra. The effect of the increased duty cycle is non-negligible internal heating effects that will, among other effects cause the lattice to expand and thus reduce energy gaps, reducing/increasing emission energy/wavelength. As is to be expected, the effect on the lasing line peak position is minimal due to the DFB structure forcing lasing to take place at the wavelength determined by the DFB waveguide’s modes. Comparing the SSL to non-SSL containing structures the addition of the SSL appears to reduce the effect of increased duty cycle on emission energy by a considerable degree. This indicates that the inclusion of the SSL produces a structure with an emission energy less sensitive to internal heating effects, a point of particular significance for the intended application of these lasers as a part of an OEIC, where processes such as wavelength division multiplexing (WDM) would require spectrally pure and stable emission. Note that a comparison of the third peaks (1350 nm) is not made since both below threshold and in pulsed operation the SNR is too small to determine the peak position with sufficient accuracy or precision.

Since the lasing line is located within this peak the implication is that this emission is from the active region as the DFB pitch would be chosen to match this emission wavelength. Here this hypothesis is further supported since common to both structures is the convergence of the second peak (triangles) below threshold with the lasing line (circles) at ~60 °C. This is significant since it is around the expected operating temperature
of a CPU or GPU when under load assuming suitable cooling. Even idling, these components would be expected to operate above ambient temperature and are designed with a thermal cut out that will activate at, at most 100 °C [39], giving a standard operating temperature range of ~30-90 °C. Thus the optical components to be integrated into these microelectronics would be expected to be designed to operate around the centre of this range, i.e. around 60 °C. Interestingly, when operated under CW injection and above threshold (here 120 mA) the peak attributed to emission from the active region is significantly redshifted, changing the convergence temperature from ~60 °C to 0 °C and ~25 °C in the case of the SSL and non-SSL containing structures respectively. Operating a complete system at these temperatures would require significant active cooling, greatly reducing overall electrical efficiency and thus these devices would not have been designed with this convergence point. Since the redshift results primarily from the internal heating caused by phonon generation and ohmic heating, through non-radiative recombination if these efficiency limiting processes were to be minimised the convergence point would return to that observed in the case of the sub threshold behaviour, fitting the operating temperature range of the system as a whole.

![Figure 4.17: Wavelengths of each peak for SSL containing (left) and non-SSL containing (right) structures as a function of temperature for both injection regimes. Square data represents the large central peak, triangular data the longer wavelength satellite peak, starred data the shorter wavelength satellite peak and circular data the lasing line. Solid and hollow symbols indicate pulsed and CW injection respectively and black and blue symbols represent below and above threshold injection respectively.](image)

Maintaining a constant temperature of 20 °C, additional spectra were produced under CW injection over a current range of 10 - 100 mA (0.89 - 8.93 kAcm⁻²) in order to identify the injection current dependence of the spectral features observed in the temperature dependent study above. Figure 4.18 shows these plots for both structures, SSL containing on the left and non-SSL containing on the right, with figure 4.19 summarising the positions of each peak as a function of current for both SSL (black) and non-SSL (blue) containing structures. Spectra produced under pulsed operation are not presented here since the effect of duty cycle on the spectra has already been determined from the temperature dependent spectral analysis and the issue of low SNR, especially at low injection would make interpretation of data with a reasonable degree of confidence
As injection current is increased the filling level of carriers in the active region rises also, since the quasi-Fermi-levels split by a greater energy as detailed in section 2.1.3. Thus the general trend expected with current is a blue shift in the emission spectrum because the energy separation of the injected carriers is now larger. This effect is known as the Burstein-Moss shift [161]. The narrower the active region the larger the increase in the quasi-Fermi-level separation for a given injection current and so of these three peaks (assuming the separate emission source hypothesis) the longest wavelength peak (second peak) on which the lasing emission is centered would be expected to blue shift the largest amount since this is the emission from the quantum well. However from the summary plot presented in figure 4.19 we see that this peak instead red shifts with $\frac{\lambda}{dI} = 0.47 \text{nmA}^{-1}$, whilst all other peaks were found to red shift by, at most 1 nm over the full current range, to all intents and purposes unchanged. The fact that the lasing peak is not found to move with current is not unexpected since the pitch of the DFB waveguide structure alone determines the wavelength of this emission.

The red shift of the peak that is assumed to originate from the active region however is counter to expectations based on the Burstein-Moss shift. Two possible explanations include internal heating and bandgap renormalisation. If the net effect of one or both of these were so great that the resulting red shift more than counteracts the blue shifting from the Burstein-Moss shift then a net red shift could be observed. It was found earlier that carrier leakage and/or Auger recombination are the most likely candidates for the dominant recombination mechanism in these devices, both of which increase strongly with carrier density, and, hence injection current. Thus, energy wasted by these processes by either ohmic heating or the generation of phonons directly from injected carriers could potentially cause a strong current dependent heating effect to counteract the expected blue shift and cause the observed red shift in the active region emission. If
equation 2.2, relating the bandgap to temperature is used along with the Varshni coefficients obtained from Vurgaftman [102] the internal temperature required to achieve the observed red shift can be calculated and is found to be approximately 380 K or 107 °C. As mentioned previously high quality microelectronics can reach temperatures of ~100 °C and so a comparably resistive and inefficient laser reaching such internal temperatures is feasible. This is made all the more likely in these devices since the III-V mesa is insulated by the SiO₂ buried insulator layer, making the conduction of excess heat away from the active region slow, allowing the temperature to rise. Internal heating would also be consistent with the observed lack of pinning and has been associated with a lack of pinning in AlGaInAs QW lasers previously [68]. The Fermi-distribution of carriers is increased by the elevated thermal energy, requiring a greater injected carrier density to maintain the modal gain as carriers are increasingly excited into higher states from which gain is not generated. If the heating results from pumping and non-radiative recombination then the effect is essentially that the threshold current density is increasing with pumping, leading to a lack of pinning.

Bandgap renormalisation, as detailed in section 2.13 could also cause the observed redshift since current densities in these devices are large. The current range of 10-100 mA corresponds to a current density range of 0.89 - 8.93 mA/cm², so even at low injection the current density in the active region is significant. Thus it is reasonable to conclude that an observed red shift, if not entirely accounted for by internal heating due to the large current density could also receive a contribution due to the bandgap renormalisation effect. In either case the fundamental cause for the red shift is a large current density in the active region, so maximising device efficiency in order to minimise the threshold current will limit the red shifting of the active region.

The remaining two peaks (central and third), if originating from the emission of layers in the structure other than the quantum well could be expected to be fairly current independent. This is because carrier confinement will not have been engineered into these regions, making the Burstein-Moss or bandgap renormalisation less prominent since carriers can flow out of these other regions. The effect of internal heating is somewhat more difficult to determine as it is dependent on the dissipation of heat from the active region. It is however noted that the other layers are significantly thicker and therefore exhibiting a larger heat capacity reducing the effect of heating on the bandgap.
Figure 4.19: Wavelengths of each peak for SSL containing (left) and non-SSL containing (right) structures as a function of CW current. Square data represents the large central peak, triangular data the longer wavelength satellite peak, starred data the shorter wavelength satellite peak and circular data the lasing line.

In addition to those collected through the substrate, spectra were also produced from the emission at the facet. These were carried out at ambient temperature for injection currents of 10, 50 and 100 mA (black, red and blue respectively) and plotted in figure 4.17 below. Firstly it can be seen that below threshold (black line) a gap in the spectrum is found centered on 1592 nm, where the higher current spectra exhibit a lasing line. The sharp drop in the spectrum at this point is the result of the DFB structure reflecting back the lasing wavelength into the below transparency active region where it can be reabsorbed. At the higher currents the narrow lasing peak can be seen at 1592 nm with a smaller satellite peak either side on top of the scattered spontaneous emission from the III-V region. The modulation across the spectrum is a cavity effect, the origin of which can be found in the discussion of figure 2.12.

It can be seen that spectra produced from facet emission exhibit the same broad peak that was hypothesized to be the spontaneous emission from the active region as well as a smaller peak at ~1500 nm that is in the same position as the central peak in the substrate spectra. The signal drops to the noise floor of the detection system below ~1450 nm and so the third peak from the substrate spectra is not observed. It is believed that the weaker spontaneous emission from the “facet” (the spontaneous emission is not coming from the waveguide but the III-V layers above it) in comparison to the substrate emission spectra results from the geometry of the pumped current channel (see figure 4.1). The pumped area when viewed from above (the growth direction) is long and narrow, aligning with the waveguide, current flowing downwards towards the substrate. Thus photons produced in this pumped region traveling downward through the substrate pass
only through pumped, and therefore relatively transparent regions of the structure. In contrast emission in
the direction of the waveguide (but still within the III-V mesa), towards the facet, other than that passing
along the cavity will pass through unpumped, and therefore absorbing material before reaching the surface
of the device. Thus it is to be expected that emission other than the lasing line would be more intense when
collected through the substrate rather than the facet, as is observed here.

Figure 4.20: Facet emission spectrum from SSL containing device at 10, 50 and 100 ma under a CW injection
regime.

In order to determine the origin of the three peaks in the spectra, LaserMod from Rsoft was used to calculate
the approximate dispersion relation of the quantum wells, barriers and SCH layers using the kp method.
The calculated dispersion relations were used to plot the band alignments for the structure of the separate
confinement heterostructure as shown in figure 4.21. The composition of these layers are found in table 4.1
with details of LaserMod in section 6.1.2. Converting the minimum energy gap between the heavy hole and
conduction band dispersion relations to wavelength gives 1595 nm, 1465 nm and 1360 nm for the QW, SCH
and barriers respectively. Comparing these values to the spectra above yields a very close match indicating
that the low energy peak, on which the lasing line is centered originates from the QW, as expected. The
high energy satellite peak matches with the barrier emission wavelength and the large central peak with the
emission wavelength of the SCH layers.

The fact that the peak from the SCH is emitting more strongly than the QWs is somewhat counter
intuitive at first since this region is not confined and the carriers should have good mobility in this region.
However if the total thicknesses of the regions are compared (see table 4.1) we find that the QWs sum to
56 nm whereas the SCH layers sum to 250 nm, so the volume of the SCH material is ~5 times that of the
QWs. Thus spontaneous emission from the SCH layers will be enhanced relative to the quantum well simply
as a result of the considerably larger volume from which spontaneous emission can take place. The emission from the barriers being of comparable intensity to that from the QWs is logical in the sense that the volumes of these regions are similar with thicknesses of 90 nm and 56 nm respectively. However since the barriers, by definition inhibit the flow of carriers into their material and in doing so confine the carriers to the QWs one would expect significantly stronger emission intensity from the QWs. With a pumped area of $11.2 \times 10^{-6}$ cm$^2$ the current density varies from 0.89 - 8.93 $kAcm^{-2}$ (or 10 - 100 mA), with such large current densities, even at low injection it can however be summarised that the band filling could conceivably extend beyond the top of the barriers, allowing these regions to house sufficient carrier densities for significant spontaneous emission to take place. From the el level the confinement from the barriers is comparable to $4 k_b T$, and so with band filling taken into account this seems like a reasonable circumstance and since the electrons must be able to tunnel through the barriers to reach the other QWs and the confinement from the SCH is only 45 meV leakage would also be expected into this region. This again fits with observations of strong carrier leakage as the apparent dominant recombination mechanism along side Auger recombination.

The band offset between the QW HH1 and barrier HH1 edge is found to be only $\sim 29$ meV at ambient temperature, comparable to the thermal energy of carriers at this temperature of $\sim 26$ meV. Carrier leakage usually takes place via conduction band electrons escaping either into the conduction bands of surrounding layers, be that $\Gamma$ to $\Gamma$-minima or $\Gamma$ to X or L satellite minima or to deep defect levels in the surrounding materials. This is the case since free valence band holes have a larger effective mass and hence a lower mobility than free conduction electrons and therefore tend to be less likely to leak from the active region before recombining. However, it has been previously demonstrated [166] that carrier leakage through the valence bands is a significant source of loss for laser diodes based on the AlGaInAs/InP material system. This is the case because the materials used in these devices have similar band gaps, with most of the offset between the conduction bands. Thus the valence band offset is generally very small, as is the case here, allowing carrier leakage to take place through the valence band, regardless of the low hole mobility. Therefore, it is feasible that this is the case in these devices also, since aside from being bonded to silicon the active region is near identical to other AlGaInAs/InP laser diodes developed elsewhere.
Figure 4.21: The band alignments of the separate confinement heterostructure used within the AlGaInAs hybrid Si evanescent lasers calculated using the k.p. method function within LaserMod.

4.2.6 Distributed feedback AlGaInAs-silicon evanescent laser summary and conclusions

Section 4.3 details studies on Fabry-Perot AlGaInAs-silicon evanescent lasers, a different production run of the structures studied in this section but with a Fabry-Perot rather than DFB grating type. These devices allow further experiments to be carried out without the effects of the DFB waveguide having any type of masking effect on the underlying physical mechanisms limiting device efficiency. As such, the results of the DFB structure study are summarised in this subsection along with any conclusions that can be drawn as to dominant efficiency limiting processes and physical properties. These conclusions can then be easily compared to those of the following section, allowing a clearer picture of the physical properties of the AlGaInAs-silicon evanescent laser to be formed.

This study concludes that SRH, Auger recombination and carrier leakage are all significant loss mechanism in these devices, with the latter two becoming particularly limiting at increased temperatures and injection currents. Further amplifying these mechanisms above threshold is the observed non pinning, which will strongly limit the maximum operating temperature and output power. The low yield and apparent inconsistent quality across the wafer, implies that the bonding process although strong on a macroscopic level may exhibit inhomogeneous quality on a smaller scale, giving variable performance across devices diced from the wafer. The offset in carrier leakage between SSL and non-SSL containing devices appears to independent of temperature and if simply a result of only the best non-SSL containing devices lasing would be tied to the bond quality. It therefore appears that the bonding process still requires refinement if highly efficient devices
are to be produced at an acceptable yield.

The primary effect of the SSL is to reduce the defect density in the active region, which is reflected by the reduction in defect current fraction from $40 \pm 2\%$ to $29 \pm 2\%$ at $10^\circ C$. PL, TEM, Z analysis and $T_1$ all indicate a reduction in defect density in the active region through the effective blocking of defects, and as a result an increase in device efficiency through the reduction of the fractional defect current at threshold. In addition one of the dominant loss mechanisms in these devices, carrier leakage, is reduced by the inclusion of the SSL. This is thought to be a result of the removal of an additional carrier leakage path created by regions of high defect density within the active region. Although the SSL is found to increase the device yield, the devices which become operational exhibit a considerably larger threshold current and as a result are not of much use. For an increased yield of efficient as opposed to barely functional lasers, changes to the bonding procedure would likely be required.

So although the SSL is beneficial, the dominant efficiency limitations appear to be related to the bonding quality, since both structures exhibit significant SRH recombination ($29 \pm 2\%$ even with the SSL at room temperature) as well as strong carrier leakage and a lack of pinning. These are the issues that need to be focused on in order to achieve the potential of these devices, with all of these expected to be related to the quality of the bonding achieved. Auger recombination may well be brought down to acceptable levels simply through the reduction in threshold current and achievement of pinning that comes with the aforementioned improvements. However, this mechanism is to be expected at this emission energy and has been found to be the dominant source of loss in previous studies of AlGaInAs QW lasers [130].

### 4.3 Fabry-Perot cavity hybrid AlGaInAs-silicon evanescent laser

Investigation of the DFB waveguide devices in section 4.2 provided an insight into the efficiency limiting mechanisms taking place in the hybrid AlGaInAs-silicon evanescent lasers, identifying carrier leakage to be a major non-radiative recombination mechanism. Further it was indicated that the bonding processes may require refinement in order to yield reliable and consistent devices across the wafer. An issue in investigating the DFB waveguide devices was the large bars containing 50 lasers and photodetectors, which due to the difficulty in cleaving silicon, especially with tightly packed structures could not be processed into individual devices.

Thus mounting these in either the hydrostatic pressure or cryostat systems was found to be impractical preventing low temperature of high pressure characterisation from be carried out. Furthermore, although the DFB waveguide has several benefits to a finished device such as removing the need for cleaved/polished facets and yielding a narrow line width emission its presence makes identifying the underlying physical limitations of the devices more challenging. The issue being that the lasing line is tied to the DFB pitch rather than the band structure, limiting the ease with which information on this can be inferred. Furthermore parameters such as the absorption coefficient, easily determined from relation 2.38, assuming a Fabry-Perot cavity for example also become difficult to study. As such it was requested that a batch of devices utilising a Fabry-Perot cavity based on the most recent device processing iteration be provided for the continuation of the
The devices received consisted of a single die containing hybrid AlGaInAs-silicon evanescent lasers of varying length: 200, 400, 600, 800 and 1000 µm. The structure of the III-V region is unchanged from that detailed previously in Table 4.1. The SOI region contains a waveguide of width 1.5 µm and height 0.76 µm consisting entirely of silicon, etched according to the same process as detailed at the beginning of section 4 with the omission of the patterning of the grating structure. Post bonding the wafer was diced to form the section we have received and the facets polished to form basic, un-coated facets in order to complete a cavity with a confinement factor of 5% with the QWs, close to that achieved in the DFB waveguide devices.

4.3.1 Ambient temperature characterisation

Initially the room temperature investigation detailed in section 4.2 for the DFB waveguide devices was repeated for the Fabry-Perot devices in order to allow a quick comparison of the key properties of the lasers; threshold current density, differential quantum efficiency etc. These were repeated using the same probe station system and pulsed injection scheme (frequency of 10 kHz and pulse width of 500 ns) in order to minimise the number of variables which may cause uncertainty in the comparison. Electroluminescence was carried out on each of the devices on the die and it was found that all were operational. Having only a sample size of 5 it is not really possible to make a judgment of yield as was the case with the two bars of 50 devices from the DFB investigation as it is likely that such a small sample would have been selected to carefully contain exclusively operational devices.

LI curves over the temperature range 283 - 323 K are depicted in the left plot of Figure 4.22 for the device of length 1000 µm. Over the investigated temperature range the threshold current density was found to vary from 5.36 to 10.18 kAcm⁻², with a value of 6.16 kAcm⁻² at ambient temperature. In comparison the SSL containing DFB results over the same temperature range were 4.64 to 10.63 kAcm⁻², with a value of 5.27 kAcm⁻² at ambient temperature. At lower temperatures the DFB waveguide device is seen to exhibit the lower threshold but by 323 K the Fabry-Perot device has a marginally lower threshold. The DFB cavity would be expected to exhibit a lower threshold current density than the Fabry-Perot cavity since modes other than the lasing line do not undergo reflection and as such are not consuming carriers that could otherwise contribute to gain. This of course assumes that the DFB pitch has been chosen appropriately to align with the peak of the gain spectrum at the expected operating current and temperature. However, because the bonding process has been further optimised since the DFB containing devices were produced a reduction in non-radiative recombination would be expected, lowering the threshold. It was indicated in section 4.2 that carrier leakage and Auger recombination are dominant recombination mechanisms in these devices and so it is logical that the Fabry-Perot device would perform worse until larger temperatures where Auger recombination and carrier leakage account for such a large fraction of injected carriers that the benefits of the DFB cavity are canceled out in terms of threshold current density. This is illustrated with the aid of the right hand plot of Figure 4.22, where the rate of increase of threshold current density for the DFB cavity device can be seen to increase at a greater rate than the Fabry-Perot device with temperature, with a $J_{th}$
cross over at \( \sim 320 \) K.

The differential quantum efficiency in contrast to the threshold current density is consistently superior in the case of the Fabry-Perot cavity device with a value of 31%, compared to 21% at ambient temperature, dropping to 26% compared to 6% by 323 K. This is again consistent with the optimisation of the bonding process leading to a reduction in carrier leakage and therefore increasing efficiency. The slope efficiency, as given by equations 2.34 and 2.29 is dependent on the mirror losses of the cavity and is only correct for a Fabry-Perot cavity since for a DFB structure the photons are undergoing many reflections when traveling any given distance as oppose to two per round trip of the cavity for a FP laser. Thus when comparing the slope efficiencies of the DFB and FP cavity lasers it is important to take into account that a difference in cavity losses will affect the measured efficiencies but to a degree that is not easily quantifiable. Therefore the comparison of these numbers should be taken as a rough trend, indicative of potential changes in recombination mechanism fractions but not as absolute values.

Also of note is how much more stable the efficiency is with injection current, again (assuming a non-pinning) indicative of a reduction in carrier leakage and Auger recombination, which would be expected to increase with band filling. The injection current was raised up 35 \( kAcm^{-2} \) in the investigation of the Fabry-Perot cavity devices compared to only 13.5 \( kAcm^{-2} \) in the case of the DFB cavity devices, since in general these were found to become unstable or damaged at current densities approaching 20 \( kAcm^{-2} \). The resilience to increased currents (35 \( kAcm^{-2} \) corresponds to \( \sim 400 \) mA injection current passing through the device) is again indicative of improved bonding since the current channel passes close to the bonding region and so if material quality is greater and more consistent the material will be less resistive and therefore less likely to be damaged by the large current densities.

![Figure 4.22](image)

**Figure 4.22**: Left: The I-I curves for the 1000 \( \mu m \) Fabry-Perot waveguide device over the temperature range 283 - 323 K. Right: A comparison of the Fabry-Perot cavity threshold current densities to those of the SSL containing DFB structure from section 4.2.
The availability of devices with a range of cavity lengths allows investigation of the internal loss coefficient and internal quantum efficiency to be carried out using the relations found in section 2.8. The internal quantum efficiency, as defined by equation 2.36 represents the efficiency with which injected carriers are converted to stimulated emission photons within the active region, unlike the external quantum efficiency which is the efficiency with which injected carriers are converted to stimulated emission photons emitted from the cavity. The relevance of the internal quantum efficiency then, is that it is more directly related to the physical properties and efficiency limiting processes of the material system than the external quantum efficiency, which is reduced by additional optical losses. According to equation 2.37 if both of these efficiencies are known, in addition to the mirror losses (which can be calculated from the refractive index contrast of the mirrors (facets) and surroundings) the internal optical loss can be determined, which in turn is representative of the presence of non-radiative recombination sites, intervalence band absorption, free carrier absorption, scattering etc. If one considers equation 2.38 it can be seen that as the cavity length tends to zero $\frac{1}{\eta_d}$ tends to $\frac{1}{\eta_i}$ and therefore in that unphysical limit they would be equal. The relationship between $\frac{1}{\eta_d}$ and $L$ is of the form of $y = mx + c$ however, making the intercept of a plot of $\frac{1}{\eta_d}$ vs $L$ with the y-axis equal to $\frac{1}{\eta_i}$, which can be determined through extrapolation of the linear dependence back to a zero value if slope efficiencies for multiple cavity lengths are known.

The intercept of the inverse slope efficiency $(1/\eta_d)$ determined from the LI plots extrapolated to 0 cavity length with the y-axis of figure 4.23 gives the internal quantum efficiency and the internal loss is determined from the gradient. The lowest cavity length data point is considered an outlier due to a deviation from the linear relationship predicted from equation 2.38 at short cavity length. At ambient temperature the internal quantum efficiency was determined to be 69 ± 4% and the internal loss 19±2.2 cm$^{-1}$. The internal quantum efficiency is less than has been previously reported for this material system at 69 ± 4% in comparison to ~80% [141]. Since the internal quantum efficiency is limited by non-radiative recombination the indication is that the degree to which these processes is taking place is therefore greater than previously reported devices of the same material system. Since the pre-bonded III-V wafer is of good material quality the implication is that the additional non-radiative recombination results from perturbations to the III-V material caused during the bonding process.

The internal loss of 19±2 cm$^{-1}$ is somewhat greater than the 15 cm$^{-1}$ of the AlGaInAs QW devices reported in [141], indicating that optical losses in these devices may be a contributor to the lower external quantum efficiency. In comparison to industry standard InGaAsP/InP lasers, where internal losses of 1.3 cm$^{-1}$ have been reported [38], the value is far larger. This suggests that although non-radiative recombination appears to be the main limiting mechanism in these devices there is still a lot of scope for reduction in optical losses to be made in order to further optimise device performance to the levels achievable in commercial communications grade lasers. This is potentially related to the evanescent coupling of the mode from the waveguide into the active region. Previous studies into 1.5 µm InGaAs(P) MQW lasers [121] found IVBA to be a prominent source of optical loss, with values of ~10 cm$^{-1}$ measured indicating that it is probable that IVBA is also taking place to a similar extent in the hybrid devices studied here. It may be that the geometry
of these lasers, by which the mode is evanescently coupled through the III-V into the SOI waveguide is at least partly responsible for the increased optical losses.

![Graph](image)

**Figure 4.23:** Inverse slope efficiency plotted vs cavity length from which internal quantum efficiency is determined from the intercept and internal loss from the gradient.

The spontaneous emission, as in section 4.2.2 required correction by accounting for scattered stimulated emission using linear interpolation. Doing so yielded the plots presented in figure 4.24 below. The most important point of note here is that the pinning is considerably better in these devices when compared to the DFB waveguide lasers. In the DFB waveguide devices at the lower temperatures a reasonably sharp pinning was observed initially, which degraded with increased injection current. However, in the case of the FP devices we see a sharp discontinuity in the spontaneous emission at threshold and a constant gradient thereafter at all temperatures investigated here. As was discussed previously in section 4.2 non pinning leads to an increase in carrier density above threshold, which increases the proportion of injection current accounted for by recombination mechanisms with a carrier density dependence which is greater than quadratic, namely Auger recombination and carrier leakage. Therefore, since not only is pinning taking place to some degree at all temperatures it is not decaying with current, suggesting an improved internal/external quantum efficiency with current, as is observed in 4.23.

Pinning takes place because the photon density is great enough to de-populate filled energy states of carriers as quickly as they can be injected. If intervalence band absorption or free carrier absorption take place then these lasing photons are reabsorbed, and if these optical losses are great enough the photon density within the active region will be significantly decreased. Optical losses of this type have been indicated to
be taking place in the FP devices, although this could not be shown in the case of the DFB devices as a result of the non-Fabry-Perot cavity making equation 2.38 inapplicable. However, since these two devices are broadly similar and the FP cavity devices are an improvement it is probable that these optical losses be taking place to a similar, if not greater degree in the DFB containing devices. As such, a reduction in optical losses achieved by the optimisation of the bonding process may be the cause of the improved performance of the FP devices through improved pinning. Previously internal heating was suggested as a potential cause of the non-pinning, however the threshold current densities are broadly similar in these devices and the device structure, other than the cavity is identical. Thus one would expect internal heating to a comparable degree in both cases and for the pinning to degrade with temperature, as was the case with the DFB cavity devices. Thus it may be that strong optical losses, reduced through improved bonding procedures are at least partially responsible for the lack of pinning in both cases.

Figure 4.24: The spontaneous emission collected from the base of each structure as a function of current over the temperature range 283 - 323 K, after correction by removal of scattered stimulated emission.

Spontaneous emission from the Fabry-Perot cavity devices was found to be particularly weak, even under CW operation, with a SNR approaching one at higher temperatures (see figure 4.25). A first point of note from figure 4.25 is that only a single peak is visible, identified as originating from the SCH layers, in contrast three were observed in section 4.2.5 associated with the QWs, barriers and SCH layers. If a single peak is found it would be expected that this were that of the QWs, however if the lasing line wavelength is overlaid (red) then it can be seen that these do not align, as they would for a FP cavity. In section 4.2.5 the strongest emission originating from the SCH layers was justified by the fact that the total volume of these layers was a factor of \(~5\) greater than that of the QWs. In the case of the Fabry-Perot lasers, since the external quantum efficiency is greater and pinning takes place to a greater degree it would be expected that the emission from the QWs be comparatively larger. The overall emission is weaker which may indicate a reduced photon extraction
efficiency from the bottom surface of the die, although compositionally unchanged the substrate is twice as thick for the FP cavity devices. Thus absorption is unlikely the cause since both Si and SiO₂ exhibit high transmittance in this wavelength range. It is however possible that spontaneous emission, emitted in random directions from the active region is scattered to such a degree that the intensity of photons exiting directly downward is minimal. If this is the case then it is possible that only the strongest emission, i.e. that from the SCH region would be detected. The rate of change of peak emission wavelength with temperature, \( \frac{d\lambda}{dT} \) was found to be 0.38 \( nm K^{-1} \), consistent with the value of \( \frac{d\lambda}{dT} = 0.34 \) \( nm K^{-1} \) in the case of the DFB waveguide devices.

The injection current was raised from 30mA to 300mA, exceeding the lasing threshold and the spectra were repeated over the same temperature range of 283 - 323 K. The position of the SCH peak is fairly consistent with that of the lower injection peak, with a small red shift as would be expected for increased temperature. The spontaneous emission from neither the barriers nor the QWs is visible once more, presumably for the same reasoning as stated above figure 4.24. The higher energy emission from the SCH should undergo the strongest absorption due to it meeting the requirements for absorption in more layers than that lasing emission. This may not be too significant though since, from table 4.1 it can be seen that the active region is only 150 nm from the wave guide, with the bulk of that material being InP with a band gap of 1.34 eV, too great for absorption to take place. The scattered lasing emission can be seen clearly for all temperatures, red shifting strongly with temperature as would be expected since the band gap will be decreasing as the lattice expands. The rate of change of peak emission wavelength with temperature, \( \frac{d\lambda}{dT} \) was found to be 0.85

![Spontaneous emission spectrum as a function of temperature below threshold, injection current = 30mA. A 10 pt smoothing function has been applied due to the low SNR. The lasing line wavelengths are marked in red for each temperature.](image)

Figure 4.25: Spontaneous emission spectrum as a function of temperature below threshold, injection current = 30mA. A 10 pt smoothing function has been applied due to the low SNR. The lasing line wavelengths are marked in red for each temperature.
$nm K^{-1}$, consistent with the value of $\frac{d\lambda}{dT} \approx 1 \ nm K^{-1}$ in the case of the QW spontaneous emission from the DFB waveguide devices.

A point of some interest is that at 313 K two lasing peaks are observed, this represents the laser multimodalising as two cavity modes tune to an approximately equal degree with the gain spectrum. As the temperature continues to increase the laser undergoes a “mode hop”, as the original cavity mode detunes from the gain spectrum peak, returning the laser to single mode operation once more. The cavity mode which is tuned with peak gain varies with temperature since as well as the band gap decreasing the cavity is expanding, causing both the gain spectrum and the allowed cavity modes to shift. Full details of lasing modes can be found in section 2.6.1. The strong temperature dependence of the lasing peak in comparison to the DFB cavity devices as well as mode hopping indicates one of the main benefits of using a DFB cavity for wavelength multiplexed communications. If the laser temperature varies either through internal heating effects or heat transfer from the microelectronics it is integrated with, then its emission wavelength will vary comparatively strongly in FP cavity device and may even multi-mode or mode hop potentially leading to a signal that cannot be correctly demultiplexed and interpreted after transmission. In contrast the DFB emission is set by the pitch of the grating structure, which will only vary according to the thermal expansion of the layers and weak change in refractive index, giving a $\frac{d\lambda}{dT}$ of the lasing line of $\sim 0.1 \ nm K^{-1}$, as seen in section 4.2. A potential limitation that must be taken into account however, is that detuning of the active region and DFB waveguide could become problematic in terms of maximising gain.

![Figure 4.26: Spontaneous emission spectrum as a function of temperature above threshold, injection current = 300mA](image)

Room temperature investigation indicates that the optimisation of the bonding process has generally im-
proved device performance, increasing external quantum efficiency and improving pinning behaviour. These improvements appear to be the result of decreased optical losses, which in turn may be the cause of the improved pinning behaviour and as a result reduced carrier leakage and Auger recombination at higher temperatures and injection currents. Further analysis based on these results such as characteristic temperature, Z analysis etc is included into the low temperature characterisation section.

4.3.2 Fabry-Perot cavity AlGaInAs-silicon evanescent laser low temperature characterisation

Low temperature electroluminescence in essence is identical to the room temperature investigation on the probe station system however at cryogenic temperature using the closed cycle helium cold finger cryostat system (details in section 3.3). The benefit of extending the investigation to cryogenic temperatures is firstly an expansion of the temperature dependent data in order to allow trends to be more accurately identified and secondly to narrow the carrier distribution with decreasing temperature. Thus, the non-radiative mechanisms Auger recombination and carrier leakage will be reduced, as will the optical loss mechanisms IVBA and FCA, increasing device efficiency. As was the case in the room temperature investigation, pulsed injection at a frequency of 10 kHz with a pulse width of 500 ns is used in order to minimise internal heating. The temperature was varied from 300 - 40 K in steps of 40 K, first reducing the temperature and then increasing it once more in 40 K steps off set by 20 K in order to fill the gaps in the data set and identify any hysteresis or irreversible effects that may have taken place.

The LI plots are presented in figure 4.27, with the first clear observation being that the device only functioned down to a temperature of 220 K. Because non-radiative recombination mechanisms and optical losses all increase with temperature for a given structure, none of these processes could be expected to be the cause of this limit. It is possible that carriers can become “frozen” into their dopant levels if the energy required to ionise them is greater than $k_B T$, removing the majority of the free carriers in the structure and as such preventing operation. However, given that the lower temperature limit is 220 K a complete freeze-out is clearly not the case here as the thermal energy is only $\sim 2/3$ of that at ambient temperature. This would also be reflected in the IV characteristics, with no significant changes observed.

The other factor varying with temperature is the band structure and since each material has a different thermal expansion coefficient and band gap temperature sensitivity the alignment of the energy levels within the structure will vary. If the offsets of the energy levels of neighbouring layers (QWs, barriers and SCH) were reduced then a fall in carrier confinement could lead to increased carrier leakage and prevent lasing. If this were the case though one would expect an increase in threshold current, rather than a reduction and so this is not the case here either.

The threshold current density is found to fall to $2.74 \text{ kAcm}^{-2}$ per device or $0.343 \text{ kAcm}^{-2}$ per QW by 200 K, compared to $0.875 \text{ kAcm}^{-2}$ per QW at ambient temperature. The improvement in threshold current density with temperature, although significant should be considered in relation to other AlGaInAs/InP devices which have not been wafer bonded such as those presented in [152]. These unbonded devices make use of the same material system, but are found to have threshold current densities of only $0.2 \text{ kAcm}^{-2}$ per QW at
ambient temperature, exhibiting superior performance to the bonded devices investigated here at cryogenic temperatures. The temperature dependence of the threshold current is discussed in relation to figure 4.29 in terms of the characteristic temperature, $T_0$ and similarly for the external quantum efficiency and $T_1$.

The external differential quantum efficiency is found to increase from 29% at 300 K to 64% at 220 K for a 1 mm cavity length device under pulsed operation. At all temperatures the LI plots retain their peak slope efficiency up to the maximum investigated current density of $\sim 13 \text{kAcm}^{-2}$. The indication is that the efficiency limiting processes which are reduced with temperature are not strongly current dependent. This seems to contradict the previous conclusions of carrier leakage being the dominant non-radiative recombination mechanism as this would be expected to increase rapidly with carrier density. This change may however be a result of the improved pinning, which will reduce the increase in carrier density beyond threshold and therefore limit the sensitivity of the slope efficiency to highly carrier density dependent non-radiative recombination mechanisms. Optical losses in the form of IVBA would be expected to be both only weakly temperature [35] and carrier density [84] dependent and so could fit with the observation of a carrier density independent slope efficiency.

Figure 4.27: The L-I curves for the 1000 $\mu$m Fabry-Perot waveguide device over the temperature range 200 - 300 K.

Analysis of the spontaneous emission, presented in figure 4.28 indicates a good pinning at ambient temperature, as was observed in section 4.3.1 (note ambient investigation reached 400 mA injection compared to 140 mA here). As temperature was reduced it can be seen that pinning gradually degrades, counter to expectation. Optical losses would be expected to limit pinning since the stimulated photon density, which pins the carrier density is reduced, however these would be expected to decrease at lower temperatures rather than
increase. The same can be said for internal heating, the other proposed cause for the poor pinning behaviour in the DFB cavity laser investigation. As the mechanism which causes the device to cease operating below 220 K has not been identified it may well be this which is causing the degradation in pinning. Because the results were collected both for decreasing and increasing temperature a gradual change in collection efficiency with time it not consistent with observations, suggesting that the effect is a related to the properties of the device its self rather than an external factor. In addition as the threshold current density decreases with temperature and spontaneous emission intensity pins at threshold it would be expected that the threshold spontaneous emission would decrease with temperature also, counter to observations. The relatively strong emission from the SCH region could contribute to this, since this emission will not pin and could also be related to the apparent degradation in pinning above threshold.

![Figure 4.28](image_url): The spontaneous emission collected from the base of each structure as a function of current over the temperature range 220 - 300 K, after correction by removal of scattered stimulated emission.

The temperature sensitivity of the threshold current and slope efficiency are presented in the form of the characteristic temperatures, \( T_0 \) and \( T_1 \). As is described in section 2.10 the characteristic temperatures are determined from the natural logarithms of the threshold current and slope efficiencies as a function of temperature and are presented in figure 4.29 below. The characteristic values for defect, radiative, Auger and carrier leakage dominated devices at 300 K are approximately \( T_0 \approx \frac{2T}{3} = 200 \) K, \( T_0 = T = 300 \), \( T_0 = \frac{3T}{3+ \frac{E_B}{kT}} \approx 50 \) K and \( T_0 = \frac{T}{(E_B/kT)} \approx 110 \) K respectively. The value for carrier leakage involves the approximation of leakage from the QW e1 level to the SCH (energies from figure 4.21) assuming no additional band filling, which at these current densities is somewhat unrealistic. \( T_0 \) is found to be equal to 91±9 K, almost twice the value calculated in the case of the DFB waveguide devices of 48±20 K. Since carrier leakage and Auger recombination have been determined to be dominant loss mechanism within these devices one would assume that the improved characteristic temperature is the result of a decrease in one or both of these
mechanisms.

The value of 91±9 K is further from the pure Auger recombination limit of 50 K at an operating temperature of 300 K, which would imply that Auger recombination may have been reduced in these devices. It is however far closer to the carrier leakage limit, which is expected to be an under estimate due to the aforementioned approximations, making the values potentially even closer. This could imply either a movement towards a leakage dominated device or the combined effect of a reduction in Auger and SRH recombination in addition to an increase in radiative recombination. The latter seems more likely since the bonding of the FP devices has been improved and at the high temperature limit the thresholds are lower. However the thresholds are actually higher for the FP devices at ambient temperature which suggests that the change in recombination mechanism fractions is a somewhat complex combination of all four mechanisms rather than an easily identifiable change in one or two.

Another factor that complicates the comparison is the de-tuning effect with temperature and injection current, the DFB mode will align well with the gain peak at a pre-determined injection current and temperature combination but for other values will not. Thus for a given injection current the gain is lower than the peak value if there is a misalignment, leading to an increased threshold current density requirement to reach threshold gain. This adds another degree of temperature sensitivity to the threshold current density, which therefore causes a variation in $T_0$.

The temperature sensitivity of the slope efficiency, quantified by the parameter $T_1$ was found to be 112±12 K, only 4 K greater than the value for the DFB waveguide devices, well within the uncertainty on the measurement. The difference, however is that in the case of the DFB waveguide devices, beyond 300 K $T_1$ drops sharply to only 18±21 K (see figure 4.11), whereas the FP waveguide devices maintain their $T_1$ value. Thus, the temperature dependent loss mechanism that activated at $\sim$30 °C in the DFB waveguide devices appears to no longer to be as large or has had its activation energy increased beyond the maximum investigated temperature of 323 K / 50 °C. In addition by comparing figures 4.5 and 4.27 it can be seen that the current dependence of the slope efficiency is also superior in the case of the FP device, again maintaining the peak value, unlike the DFB waveguide devices which exhibited a comparatively strong drop off in slope efficiency with injection current. The implication of these results is, again that temperature dependent non-radiative recombination mechanisms have been minimised through bonding process optimisation. Again the de-tuning effect will cause some changes to $T_1$ as a result of effectively increasing the threshold carrier density as a function temperature and therefore allowing the proportion of Auger recombination and carrier leakage to increase, lowering the slope efficiency.
Figure 4.29: The $T_0$ and $T_1$ variables plotted as a function of temperature from 220 - 300 K for the 1000 µm Fabry-Perot waveguide device.

$Z$ analysis, as detailed in section 2.9 and previously applied to the DFB waveguide devices in section 4.2.4 quantifies the average carrier density power-law dependence of the combined recombination from all mechanisms, the data for which is presented in figure 4.30. Half threshold $Z$ values are found to vary from 1.95 at 220 K to 2.40 at 323 K. Both values imply the presence of a temperature dependent loss mechanism, either Auger recombination or carrier leakage or some combination of the two; which in addition to SRH recombination balance to give these values rather than the devices being radiatively dominated. Once more a fit to the $Z$ values can be used to show that values close to 2 do not necessarily indicate a radiatively dominated device when performance is non-optimal, as is the case here. For details see the discussion of figure 4.13 from the DFB cavity laser study and section 2.9.2 of chapter 2.
The approximations of the SRH fractions at threshold (34±2% and 20±2% at 220 K and 323 K respectively) are added to figure 4.31 in addition to the possible fractional combinations of SRH, radiative and Auger recombination to act as a lower limit for the defect fraction at low injection (the data from which these defect fractions are extracted is included in appendix A). Once more both values can be fitted for a wide range of values other than a radiative dominated device, and with the lower limits for SRH recombination the maximum radiative fractions and minimum Auger fractions are found to be ∼37% and 34% at 220 K and ∼20% and 60% at 323 K. These are of course approximate limiting values and could differ somewhat, but do give an approximate idea of the levels of each mechanism taking place. It is clear that even at 200 K radiative recombination does not dominate and that as expected the Auger fraction is considerably lower than it is at 323 K. Although carrier leakage is assumed to be minimal as a result of the reduced injection levels if it is a prominent recombination mechanism at threshold it is likely taking place to some small extent here. As such it is stressed that these values act as an approximate guide only but certainly show that the values close to 2 do not represent a radiatively dominated regime.

Figure 4.30: Z values as a function of temperature calculated at both threshold injection current and half this value, both data sets are fitted with exponential functions.
Figure 4.31: Possible combinations of SRH, radiative and Auger recombination that can be used to fit the $Z$ values close to 2 seen at low injection, suggesting that these values do not represent a radiatively dominated device.

The trend in $Z$ value with temperature was fitted with an exponential function, which was considered valid in this case based on previous findings of carrier leakage (exponentially dependent on temperature) being a dominant source of loss along with Auger recombination. Over the temperature range of the probe station investigation the half threshold $Z$ value was found to vary from 2.16 to 2.40 in comparison to 2.18 to 2.46 in the case of the DFB waveguide devices. Since the difference between these values are negligible it is indicated that any improvements yielded from bonding process optimisation has only had an effect on processes with a strong carrier density dependence or a balanced combination of Auger and SRH.

At threshold the $Z$ value varies from 2.28 at 220 K to 3.46 at 323 K. The low temperature value of 2.28 indicates a comparatively larger radiative fraction than at higher temperatures but will still be dominated by a combination of Auger recombination, carrier leakage and SRH recombination, as is indicated from the discussion of figure 4.31. By 323 K the $Z$ value rises to 3.46, indicating firstly a drop in the radiative recombination fraction and secondly an increase in carrier leakage since the value exceeds 3 but likely Auger recombination also. As was the case with the DFB waveguide devices it is not possible to definitively state to what extent Auger recombination is taking place since it represents an intermediate value of $Z$ between that of radiative recombination and carrier leakage. The threshold $Z$ values were also fitted with an exponential function, this was found to exhibit a good fit to the experimental data without the ambiguity of the low injection fitting (where carrier leakage should be more minimal). The exponential relationship again implies that carrier leakage is the cause of the increase in $Z$ beyond the ideal value of 2 and the fact that it does not
level out or even slow in growth beyond 3 is a further indication that the impact of Auger recombination, whilst present in these devices may be less than that of carrier leakage at threshold injection and beyond.

Over the temperature range investigated using the probe station $Z$ is found to vary from 2.82 to 3.46 for the FP cavity devices in comparison to 3.47 to 4.16 for those with a DFB cavity. The rate of increase over this temperature range is almost identical with the difference in absolute values coming from an almost constant offset of $\sim 0.65 - 0.7$. It is found by comparison of figures 4.4 and 4.26 that the threshold current densities are comparable between the DFB and FP waveguide devices, and since the structures are otherwise the same it is reasonable to conclude that this would yield threshold carrier densities which are also comparable. Thus, if as indicated carrier leakage is the dominant efficiency limiting process in these devices it is logical that if the carrier densities and by extension band filling is comparable, that the rate of increase of carrier leakage and therefore $Z$ value would be similar, as is observed here. The offset in $Z$ value indicates a fairly temperature independent decrease in carrier leakage, which is somewhat counter intuitive if a single leakage path is considered as it would be expected that if the amount of leakage decreased the effective barrier height would be greater and therefore the rate of increase with temperature would be lower. If an additional leakage path is considered on the other hand a temperature independent decrease could be justified as the elimination of one of these paths, leaving only one, common to the two device structures, which would exhibit the same temperature dependence. In the context of these devices if optimised bonding leads to a reduction in bonding defects, which as mentioned previously have been demonstrated to have the potential to act as leakage paths [94], then this could fit the hypothesis of a second leakage path which has been eliminated.

Summary table 4.3 brings together the threshold currents and slope efficiencies for the FP cavity devices as a function of temperature in addition to the defect currents and fractions at threshold. Firstly, comparing the ambient values to those of the DFB waveguide structure a small decrease over the range 323 to 283 K is observed from $29 - 23 \pm 2 \%$ to $24 - 20 \pm 2 \%$. This decrease is consistent with an optimisation of the bonding process, leading to a reduction in the density of bonding defects which propagate to the active region. The defect current percentage at threshold continues to increase down to 220 K where it reaches 34%. Thus, if the dominant loss mechanism at threshold is carrier leakage, which is particularly temperature sensitive, then by 220 K the degree of carrier leakage (and Auger recombination) should have decreased significantly. Therefore the 34% of the injection current at threshold attributed to SRH recombination implies that the defect density within the active region is still reasonably large. With the SSL and optimised bonding it may be that there is an additional source of monomolecular recombination centres within the active region beyond bonding related defects. Were this the case then considerable improvements in performance could be achieved pre-bonding by improving the intrinsic material quality. It should be added that once more this analysis is somewhat affected by the potential de-tuning of the DFB grating from the gain peak, which is not an issue for the FP lasers. As a result it is somewhat difficult to make definitive conclusions or quantifications that are more than probable estimates, but which do however point towards potential lines of action that could lead to improvements.
Spontaneous emission spectra were produced over the temperature range 220 - 300 K from the substrate of the devices in the manner described in section 3.2.2 of chapter 3 and plotted as a function of temperature in figure 4.32 below. It should be initially noted that the peaks observed in these spectra are attributed to emission from the SCH region as opposed to the QWs and as such do not have any relation to the gain curve of the active region. The change in shape of the spectra results from the decreasing temperature causing a contraction in the lattice, resulting in an increase in the band gap and therefore a blue shift in the emission. Further to this the reduction in thermal spread of free carriers though the change in the Fermi-Dirac distribution narrows the emission linewidth. The temperature dependent blue shift of the peak position is summarised in figure 4.34 along with that of emission from the facet. The Facet spectra, plotted separately in figure 4.33 are also included in this plot to indicate the convergence of the two sets of spectra as a function of temperature. This convergence behaviour is discussed further with respect to figure 4.34.
Figure 4.32: Spontaneous emission spectrum at a constant injection current of 20 mA as a function of temperature from 220 to 300 K for the 1000 µm Fabry-Perot waveguide device.

The corresponding spectra produced from the facet emission are presented in figure 4.33 and have been normalised to allow a clearer comparison of the subsidiary peaks. The temperature dependence of the lasing peak position is plotted in figure 4.34 along with the associated data for the spontaneous emission spectra. One of the characteristic differences between a DFB and FP cavity is illustrated here by the additional lasing lines produced as many cavity modes have sufficient gain (see figure 2.12). At both 300 K and 280 K the laser is multimoding (multiple cavity modes lasing with comparable gain), an additional source of inefficiency since carriers must be shared between two competing processes. A single mode becomes dominant once more below 280 K in what is known as a mode hop. As the gain curve blue shifts shorter wavelength cavity modes tune towards peak gain, thus the pattern that repeats with temperature is single mode (one mode aligns with peak gain) -> multimode (two or more modes have comparable gain) -> single mode (a different single mode is not aligned with peak gain). The process of mode hopping can be problematic for wavelength sensitive applications (WDM), particularly if operating around the point of multimoding. Fluctuations in temperature and carrier density can cause the laser to “hop” between two adjacent cavity modes during operation, which can be difficult to suppress entirely.
Figure 4.33: Facet emission spectrum at a constant injection current of 100 mA as a function of temperature from 220 to 300 K for the 1000 $\mu$m Fabry-Perot waveguide device. (Current channel and waveguide width = 2.5 $\mu$m)

Combining the peaks of the emission spectra from the substrate as well the facet as a function of temperature and converting to energy yields the plot depicted in figure 4.34 below. The data sets for the QW (hollow black) and SCH (hollow blue) are found to converge as the temperature is reduced, as was observed in figure 4.32. Since the composition of the QWs, barriers and SCH layers are similar (see table 4.1) it would be expected that the bandgap would vary at a correspondingly similar rate with temperature. The theoretical variation with temperature was determined using the Varshni relation between band gap and temperature (equation 2.2) as is described in section 2.1.1, with the values for the Varshni coefficients of the quarternary materials calculated from the relevant binaries. The binary values in turn were taken from Vurgaftman’s III-V parameter review [102]. Plotting the theoretical temperature dependence alongside the experimentally determined emission peak energies shows that the QW emission follows the expected theoretical trend well. The emission attributed to the SCH however is found to deviate strongly from the theoretical temperature dependence, converging on the QW emission data. Since the spectra are weak it may be the case that the emission attributed to the SCH alone is in fact the net emission from both the SCH and QWs. At ambient temperature, where the SCH may exhibit a comparatively stronger emission the peak is likely mostly SCH emission. However as the relative intensity of the QW emission increases with decreasing temperature the peak value may simply be the mid point between the two individual peaks, eventually tending towards that of the QW at the lowest temperatures.
Figure 4.34: Spontaneous emission and lasing emission peak positions as a function of temperature as well as the QW, SCH and barrier bandgaps as calculated according to Varshni (see equation 2.2). Simulated data points are added (green triangles) to show how a combination of two peaks of varying relative amplitude may cause the observed trend.

The hypothesis of the peak being the result of contributions from both the SCH and QW was tested by simulating Gaussians of similar shape and combining them in varying combinations (1:(0, 0.2, 0.4, 0.6, 0.8)). The plots produced from doing so are presented below in figure 4.35, where the net signal peak position can be seen to move away from the peak of the SCH Gaussian towards that of the QW as its amplitude is increased. Taking these peak positions and accounting for the trend with temperature, as given by Varshni gives the green triangular data plotted in figure 4.34. Although clearly a rough estimation that does not take into account changes in inhomogeneous broadening for example, it indicates that such a hypothesis is certainly plausible and fits with observations.
Thus it is likely that the apparent convergence of the emission from two separate layers is merely the effect of a weak emission from the device substrate masking the emission of two separate layers, rather than purely the SCH as it at first appears. One important observation that can be made from figure 4.33 is that the difference between the lasing energy of the QWs and the band gap of the SCH layers is $\sim 50$ meV at threshold. From figure 4.21 the same energy difference is found to be 72 meV assuming no band filling and so the approximate band filling in the QWs at threshold can be determined from the difference in these two values, 22 meV. The difference in effective mass of the HH and CB e1 levels causes the filling to mostly take place in the conduction band and so it can be approximated to be the full 22 meV. The confinement between the QW e1 level and SCH band edge is also shown to be only 45 meV. If the approximate band filling at threshold is taken into account the confinement would have dropped to 23 meV, essentially the value of the carrier thermal energy at 300 K ($\sim 26$ meV). Clearly this approximation is fairly loose but gives an approximate value that will always be considerably less than the $\sim 4k_BT$ of confinement required to prevent the bulk of carrier leakage. Thus the proportion of free carriers that can be excited beyond the SCH band edge and escape the active region will be considerable, leading to significant carrier leakage, fitting the indications of carrier leakage being a dominant loss mechanism in these devices. The threshold current densities have been found to be large in these devices in comparison to both other AlGaInAs/InP devices [152] and industry standard InGaAs lasers [113] making these findings consistent with the hypothesis of carrier leakage being a key efficiency limiting process in these devices.
4.4 Conclusions

The combined results of both the DFB waveguide devices, both SSL and non-SSL containing as well as the Fabry-Perot cavity devices are considered as a whole here in order to draw conclusions of the efficiency limiting processes taking place in these devices. Areas requiring further investigation will be identified and potential works suggested to expand the investigation. Finally potential solutions to the efficiency limitations will be proposed and detailed where applicable.

Extrapolation of the low injection Z analysis values to threshold yielded ambient temperature values for the threshold defect current fraction of 39±2%, 28±2% and 22±2% in the cases of the non-SSL-DFB, SSL-DFB and FP waveguide devices respectively. The addition of the SSL reduces the SRH fraction by 11%, indicating that the SSL defect blocking layer has the desired effect of inhibiting the propagation of threading dislocations, an assertion which is supported by TEM images depicting such behaviour. A further decrease of 6% is yielded as a result of the optimisation of the bonding process when producing the Fabry-Perot devices suggesting that a lower density of defects is formed during bonding.

Auger recombination is well known to be problematic for devices operating around 1550 nm and accounts for up 80% of loss in InGaAsP/InP devices [80]. It would therefore be expected that at least some degree of Auger recombination takes place in these devices. However, it has also been demonstrated that the AlGaInAs material system on InP can be particularly temperature insensitive [141], a trait that devices dominated by Auger recombination would not exhibit. Further to this finding it has been demonstrated that the Auger coefficient for AlGaInAs devices is lower than would be expected for devices operating at this wavelength [166]. Within this investigation little direct evidence has been found for Auger recombination specifically as a dominant loss mechanism. The devices are found to be highly temperature sensitive, as indicated by their low $T_0$ and $T_1$ values, consistent with Auger recombination, but also carrier leakage. The latter however, has a greater weight of evidence as a dominant loss mechanism and so this alone does not require Auger recombination be taking place. Ambient temperature Z values at threshold exceed 3, the value associated with Auger recombination and rise exponentially, suggesting that Auger recombination is not dominant, since an asymptotic approach to a value of 3 would be expected in such a case. Thus it is concluded based on both observations made during this investigation as well as published literature that Auger recombination is likely taking place, but to an unknown extent. Quantification is somewhat difficult since this process is masked by evidence of strong carrier leakage. Regardless, in devices which do not completely pin above threshold, will be operated at high temperatures (as much as 100 °C), have large threshold current densities and operate at around 1550 nm Auger recombination will limit performance to some extent and as such must be taken into account.

Carrier leakage has been indicated to be a dominant source of loss in all devices throughout the investigation. The threshold and slope efficiency were found to be highly temperature sensitive, as is represented by the $T_0$ values of 54±20, 48±20 and 91±9 K and $T_1$ values of 53/13±21, 118/18±21 and 112±12 K (non-SSL-DFB, SSL-DFB, FP). Z analysis also indicates carrier leakage, threshold values at ambient temperature were found to be 3.26, 3.6 and 3.05 in the case of non-SSL-DFB, SSL-DFB and FP cavity devices and since
values in excess of 3 can only be achieved by carrier leakage, it must be present to some extent. Aside from
the absolute values \( Z \) continues to rise exponentially above 3 rather than asymptoting, consistent with carrier
leakage.

Simulation of the band alignments using the k.p. model, in combination with measurements of the
spontaneous emission spectra led to the finding that the separation between the highest filled QW levels and
the SCH conduction band edge was only \( \sim 22 \text{ meV} \) including approximate band filling \( (\sim 45 \text{ meV without}) \).
With \( k_b T \approx 26 \text{ meV} \) the probability that carriers would have the energy to exceed the energy barrier is high,
supporting the hypothesis of carrier leakage being a dominant loss mechanism. Furthermore the valence
band offset between the QW \( HH_1 \) level and SCH VB edge is \( \sim 23 \text{ meV} \) (for unfilled bands), and it has been
demonstrated elsewhere [166] that leakage through the valence band is a significant source of loss in AlGaInAs
active region lasers.

The internal efficiency and internal optical loss to be determined to be \( 69 \pm 4\% \) and \( 19 \pm 2 \text{ cm}^{-1} \) at
ambient temperature respectively. The optical loss is larger than reported from other studies of both
AlGaInAs/InP[141] and InGaAs(P)/InP [121], where values of \( 15 \text{ cm}^{-1} \) and \( 10 \text{ cm}^{-1} \) are presented re-
spectively. In both cases IVBA is determined to be the source of the optical losses, which given the similar
material systems suggests that IVBA is an issue common to these materials as a whole. However, in compar-
ison to the industry standard devices these optical losses are \( \sim 10x \) greater, suggesting there is considerable
scope for improvement in this area. The difference in comparison to commercial lasers could also represent
increased scattering losses resulting from lesser material quality, exacerbated by the strain induced defects
resulting from integration on Si. At the same time IVBA has been demonstrated to have only a weak temper-
ature and carrier density dependence [35, 84], which is inconsistent with the low \( T_0 \) and \( T_1 \) values. This does
not discount the presence of IVBA but does indicate that it is of secondary importance to carrier leakage
and Auger recombination as an efficiency limiting process.

4.4.1 Suggestions for device improvement

The main issues with these devices appear to stem from the inherent limitations of the AlGaInAs/InP material
system, exacerbated by integration on Si, as would be expected. Carrier leakage, potentially through the
valence bands results from the compositional similarities of the quantum wells, barriers and SCH layers as
well as the InP “substrate”. If the composition of the layers is adjusted such that the carrier confinement
is maximised the refractive index contrast with the InP layer will drop reducing the optical confinement
and as a result the optical confinement factor. Carrier and optical confinement will therefore need to be
(and presumably have been) carefully adjusted to achieve a compromise which leads to maximised overall
performance. It may be possible to improve p-side carrier confinement provided the evanescent coupling
remains strong as the silicon waveguide offers a larger index contrast than the p-side InP, however this would
require simulation to verify. Short of changing the material system choice for the III-V mesa there is little
else that can be done to rectify this situation. This is especially the case if the leakage is taking place through
the valence bands where the offset change is minimal as a function of concentration. If this is the case it
does mean that optical confinement could be prioritised more strongly and so could potentially yield slight efficiency improvements thusly. Optical confinement is also further complicated in these devices however, since not only must the overlap with the active region be considered, but also coupling into the waveguide in the SOI substrate.

Assuming the III-V mesa structure is optimised and the intrinsic material quality high then any further improvements will need to take place with respect to bonding. AlGaInAs lasers on InP rather than silicon, although exhibiting the same type of efficiency limitations as the hybrid lasers do so to a lesser extent. Thus the implication is that the bonding process exacerbates the intrinsic limitations of the material system and so further optimisation is required to allow the III-V mesa to operate to its un-bonded limits. The threshold defect current indicates that defect density in the active region is non-negligible which in turn suggests that the inclusion of strain, most likely during bonding causes defects in the lattice to form. The requirement of a defect blocking layer in its self indicates that the bonding process is damaging and induces considerable strain on the lattice.
Chapter 5

5 Development and Optimisation of Monolithic GaNAsP/Si Quantum Well Lasers

5.1 Introduction and Background

5.1.1 Monolithic integration of III-Vs on Silicon as a solution to optoelectronic integration

In Chapter 1 it has been shown that bulk silicon is an unsuitable active material for the efficient generation of photons electrically; unlike the traditional III-V semiconductor laser/LED materials, used in modern optical communications systems. Hybrid integration of III-Vs on silicon through wafer bonding techniques was presented in Chapter 4 and shows promise as a potential method for the realisation of OEICs, but have the disadvantageous property of requiring an additional processing step to couple materials of intrinsically incompatible structure leading to increased cost and theoretically reduced reliability, as discussed in sections 1.2.2.1 and 1.2.2.1.

Monolithic integration of III-Vs onto silicon would be a superior alternative, however the aforementioned lattice constant and thermal expansion coefficient discrepancies make this untenable for traditional III-Vs due to the large density of defects and dislocations generated during the epitaxial growth process [124]. Thus it would be desirable to develop a material with the photon generation properties of III-V materials such as GaAs, which also exhibits both a lattice constant and thermal expansion coefficient which are close to that of silicon.

![Figure 5.1: Band gap as a function of lattice constant for silicon and III-V binaries in addition to the quarternary GaNAsP (image from [18]).](image-url)
As a result of a lattice constant mismatch of only 0.4% [77] at ambient temperature (see figure 5.1), defect-free epitaxial growth of the III-V material GaP on silicon has long been possible [78]; however since GaP is also an indirect band gap material it is also unsuitable for use as a gain medium. The addition of arsenic to form GaAsP leads to a direct band gap, however the lattice constant mismatch increases rapidly, even more so in the case of GaInP. Small concentrations of nitrogen (<5%) have been added to III-V materials to develop so called dilute nitride materials [24], which exhibit a large reduction in band gap when only a few percent of nitrogen is introduced to the alloy.

The stark changes in band structure as a result of the nitrogen atoms is well described by the band anti-crossing theory [79] and is summarised below in section 5.1.2. Adding nitrogen to GaP forms GaNP, for which the nitrogen impurity level \( E_N \) is found below both the GaP conduction band \( \Gamma \) and X-minima and therefore, as will be discussed in section 5.1.2 the lowest conduction band \( E_- \) exhibits little \( \Gamma \) – character [27]. From figure 5.2 it can be seen that GaNP can become direct gap with just a few % of nitrogen, however the band gap will be large, at 1.75 eV, corresponding to an emission wavelength of only \( \sim 710 \) nm, far from the target of 1550 nm for 5% nitrogen. Thus GaNP is not suitable for emission at this wavelength, especially when confinement within a QW will make the band gap even larger. Furthermore as can be seen from figure 5.1 the addition of nitrogen causes a reduction in lattice constant that will increase the lattice mismatch with silicon.

Figure 5.2: The conduction band edge energy for GaAsP (\( \Gamma \) and X-minima) and GaNAsP (\( \Gamma \) only) for varying As composition.

As fraction (x)

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Conduction band edge energy (eV)

0.75 1.00 1.25 1.50 1.75 2.00 2.25 2.75 3.00

\( \Gamma \)-GaAsP

\( \Gamma \)-GaNP

\( \Gamma \)-GaNAsP

\( X \)-GaAsP

Increasing N %

\( E_N \)
It is however found that by alloying GaP with a small concentration of nitrogen (<10%) in addition to a large concentration of arsenic (>70%) the dilute nitride GaNAsP can be formed, which exhibits both a direct band gap as well as closely matching the lattice constants of GaP and silicon [18]. In addition, as can be seen from figure 5.2 GaNAsP with large As concentration has a band gap of several hundred meV less than GaNP, making achieving lasing close to 1550 nm more achievable.

Defect free, pseudomorphic growth of GaNAsP on GaP as well as Si have been successfully demonstrated by the research group of Professor K. Volz and Professor W. Stolz at Phillips University Marburg, Germany [17],[134]. Furthermore, lasing at 300 K on a GaP substrate with a threshold current density of \( \sim 4 \, kA/cm^2 \) [34] and up to 165 K with a threshold current density of 1.6 \( kA/cm^2 \) on a silicon substrate [66] has been observed within Professor Sweeney’s research group at the University of Surrey.

Thus the GaNAsP material system exhibits promise for use in realising OEICs on silicon. However the large threshold current density, even on a GaP substrate and cryogenic maximum operating temperature on a silicon substrate indicates that there is a considerable amount of optimisation required to achieve a commercially viable product. In order to identify these optimisations an understanding of the underlying physical processes that limit the efficiency of these devices is needed; this will be the focus of the investigation presented in this chapter.

The work presented here is part of an ongoing investigation of GaNAsP/GaP/Si quantum well lasers grown at the Phillips University, Marburg developed as a potential candidate for optoelectronic integration on silicon. The effect on device performance and efficiency limiting processes of laser structure and silicon substrate orientation are investigated as part of work carried out in collaboration with N. Hossain in sections 5.2 and 5.3. The devices studied here were produced in the same period as #16195 (the to date best performing device batch) and use the same growth techniques and basic structure.

### 5.1.2 The band anti-crossing model for dilute nitrides

The addition of even a few percent of nitrogen into III-V materials yields disproportionately large reductions in band gap when compared to conventional interpolation schemes according to Vegard’s law [129] and associated bowing parameters that would be applied for standard III-V alloys. The properties of standard III-V materials such as the Bohr radius and electronegativity are broadly similar, for example, in the case of gallium, arsenic and phosphorus these are 1.22 Å /1.8 eV, 1.24 Å /2.2 eV and 1.11Å /2.19 eV respectively [58]. In comparison the values for nitrogen are significantly different at 0.55 Å and 3.1 eV, these large differences in fundamental properties mean that the a small concentration of nitrogen introduced into a III-V lattice will cause local perturbations to the lattice properties.

In the case of nitrogen it is mainly the large electronegativity (partially resulting from the small atomic radius) that causes these perturbations by acting as a weak acceptor, strongly attracting and localising electrons. The strong localised potentials yield localised wavefunctions for the nitrogen defects within the lattice, and in accordance with the Heisenberg uncertainty principle the Bloch functions from which they are comprised are delocalised in k-space.
The energy of this nitrogen defect level with respect to the host levels of the lattice can be determined using a tight-binding Koster-Slater model [163]. The result of these properties is a nitrogen defect level that is flat in k-space and insensitive to the position of the conduction band, and, as a result insensitive to both temperature and pressure [162]. In contrast the localised states will have a strong impact on the lattice conduction band via their strong electronegative attraction of the local electrons, yielding a strong perturbation to the band structure with increasing nitrogen concentration, as was noted above.

The strong perturbing effect of the introduction of even a small concentration of nitrogen impurities into GaAs is exhibited in figure 5.2 below. Here it can be seen that the ambient temperature bandgap of GaAs is reduced from its host material value of 1.42 eV to only \( \sim 0.95 \) eV by 6\% N. When the band structure deviates from a linear interpolation scheme an experimentally determined bowing parameter is introduced to account for the discrepancy [79]. However, in the case of dilute nitrides and other so called highly mismatched alloys (HMAs), for example bismides, a bowing parameter alone is not enough to reconcile the difference. In the case of HMAs the properties of the impurity atoms are so different from those of the host atoms that the interactions caused by the introduction of these impurities must be calculated in order to determine the band structure properties of the resultant alloy.

![Figure 5.3: GaAsN band gap energy vs nitrogen concentration, taken from [157].](image)

The band anti-crossing model is accepted as one of the most suitable methods for determination of the band structure of dilute nitrides [34]. The model considers the interaction of the nitrogen atoms and host lattice thusly. Electronegative nitrogen impurities localise electrons and these localised electrons undergo an anti-crossing interaction with the delocalised states of the host lattice. Since the nitrogen atoms act as weak acceptors the impurity level is close to the conduction band (33 meV below or 180 meV above in the cases of GaP and GaAs respectively) [103] and so the interaction with the conduction band will be strong whilst there should be little impact on the valence band. Consequently the conduction band is split into two bands, \( E_+ \) and \( E_- \), with energy quantified according to equation 5.1 below.
\[ E_{\pm} = \frac{1}{2} \left( E_{CB} + E_d \right) \pm \sqrt{(E_{CB} - E_d)^2 + 4V^2x} \]  

(5.1)

Here \( E_{CB} \) is the host conduction band edge, \( E_d \) is the nitrogen impurity level energy, \( V \) is the coupling parameter quantifying the interaction strength between the host and impurity sites and is based on the difference in electronegativity of the materials in this case, finally \( x \) is the fraction of nitrogen incorporated into the alloy. Energies are quoted with respect to the energy at the valence band maximum. It should be kept in mind that this method assumes that the effect of the nitrogen impurities on the valence band is negligible.

Figure 5.3 below depicts the \( E_+ \) and \( E_- \) levels as well as the host material conduction band \((E_M)\) for the cases of the nitrogen level laying above (a) and below (b) the conduction band minimum. It can be seen that the conduction band splits either side of the nitrogen level with morphology similar to the parabolic-like host bands but with a delocalised-like nature of strength dependent on proximity to the nitrogen level. Thus it can be seen that if the nitrogen level is below the host conduction band minimum (b), as is the case for GaP then the \( E_- \) band exhibits strong delocalised nature due to the immediacy of the \( E_- \) and nitrogen levels. Conversely in the case of GaAs (5.3 (a)) the \( E_- \) band is only close to the nitrogen level at \( k \)-values distant from the \( \Gamma \)-minimum and so maintains its parabolic behaviour.

Consequently, the addition of As to GaNP to form GaNAsP decreases the effective mass and density of states for the \( E_- \) band in addition to lattice matching and reducing the band gap. This can be somewhat detrimental to the threshold current since the HH valence band has a large effective mass causing quasi-Fermi-level splitting to be asymmetric. Therefore, if the CB effective mass is large (as it is for GaNP or low As concentration GaNAsP) then the splitting can be more symmetrical and therefore closer to the bandgap at transparency than it would be for the case of dissimilar effective masses. Therefore the As concentration should ideally be only as large as it need be for lattice matching purposes and the bandgap controlled by other means, such as well width or N concentration.
Figure 5.4: The band structure of two hypothetical materials for which the nitrogen impurity level is found above (a) and below (b) the conduction band minimum. Host conduction bands were modified using equation 5.1 to give the $E_+$ and $E_-$ levels.

5.1.3 Previous characterisation of the GaNAsP material system

Initially GaNAsP was grown on GaP substrates only so as to optimise the growth conditions and devices structure as well as identify the physical properties of the material system limiting device performance. Initial characterisation at the University of Surrey was carried out by J. Chamings, and continued by N. Hossain thereafter. Electrical injection lasing was found to take place up to a maximum temperature of 80 K, with a threshold current density of $0.8 \text{kAcm}^{-2}$ for a 4% N content in the active region. An important observation was a so called “orange glow” being emitted from the device surface, as depicted in figure 5.5. This emission was found to peak at $\sim 2.1 \text{eV}$, a far greater energy than the lasing line emission of $\sim 1.4 \text{eV}$, corresponding to the ground state emission of the quantum well, yet lower in energy than the band gaps of the GaP barriers and substrate. An emission in this energy range is known to be exhibited by GaP containing nitrogen impurities [158]. Also GaNP LEDs investigated as part of the same study exhibited a similar spectrum [34] leading to the conclusion that diffusion of nitrogen from the active region into the barriers, forming energy levels through the BAC interaction was the likely cause of the emission.
Figure 5.5: The observed orange glow (1.7 - 2.3 eV) found to be emitted from the device surface (taken from [34]).

Temperature dependent analysis carried out using a similar methodology to that described in section 3.3 indicated carrier leakage to be a dominant source of loss in these devices. Further, high hydrostatic pressure analysis carried out using the system described in section 3.4 confirmed carrier leakage as a dominant loss mechanism as well as indicating that the pressure coefficient of the leakage level corresponded to GaNP with a nitrogen content of 3%. Evidence of an inhomogeneous active region was also found, as indicated by a large $T_0$ and lack of pinning behaviour above threshold. Thus the conclusion of this study was that lasing of GaNAsP/GaP was possible under cryogenic conditions, although at a large threshold proving the concept had potential. The cause of the limited maximum operating temperature and large threshold was found to be strong carrier leakage into the GaP barriers facilitated by the diffusion of the majority of the nitrogen out of the active region, effectively lowering the barrier height by forming new, lower energy GaNP levels. In doing so the active region was left inhomogeneous and containing far less nitrogen than was required for optimal performance of the structure as designed. Further details of these findings can be found within the thesis of J. Chamings [34].

The continued characterisation of GaNAsP/GaP was carried out by N. Hossain on devices that had been grown and processed based on the information provided by J. Chamings along with growth condition optimisation studies within the research group at Marburg. The devices investigated by N. Hossain were found to lase up to room temperature with a threshold current density of $4 \text{kAcm}^{-2}$ with characteristic temperatures $T_0$ and $T_1$ of 58 K and 37 K respectively. Optical losses were found to be $\sim33 \text{ cm}^{-1}$ (single QW) compared to around $1.3 \text{ cm}^{-1}$ as has been reported for InGaAsP industry grade communications lasers (triple QW) [38].

By isolating the radiative component of the threshold current density the non-radiative losses were estimated to be as much as 92%. The Z value was found to rapidly increase with temperature from 2.2 at 60 K to 3.5 at 290 K, indicating the presence of a highly temperature sensitive loss mechanism. The rapid increase
in value, going above 3 indicated that carrier leakage remained a dominant source of loss for these devices, this was confirmed by a rapidly increasing threshold with the application hydrostatic pressure.

The pressure coefficient of the barrier X-minima was found to be greater than that of the leakage path and thus leakage via this path was eliminated as a possibility. The low pressure sensitivity does however support the previous hypothesis of the leakage path involving the proposed $E_n$ sub-band induced in the barriers as a result of nitrogen diffusion from the QW. The pressure dependence of the leakage levels was again used to estimate the percentage of nitrogen leaked into the barriers and used with equation 5.1 to compute the $E_n$ and $E_\pi$ energy levels of hypothetical nitrogen containing barriers.

Thus the conclusion of the study was that although device performance had improved, to achieve room temperature lasing, major improvements were still required to reach the standard required for commercial applications. It has however been made clear that the issues of the devices, namely carrier leakage and optical absorption are both connected to the hypothesised diffusion of nitrogen from the active region into the barriers, giving a target optimisation at the growth stage to limit this process.

GaP, although exhibiting a mismatch of only +0.4% in lattice constant with silicon can still be problematic to grow in thick layers epitaxially on silicon as a result of a difference in thermal expansion coefficient of +44%. The optimum growth temperature for GaNAsP is reported to be 575 $^\circ$C [66] since at this temperature the degree of nitrogen diffusion into the barriers was calculated to be lesser than was determined for devices grown at higher temperatures. Thus, at growth temperature the lattice constant mismatch is greater than at ambient temperature, resulting in significant strain build up within the lattice. Since defects and dislocations are primarily induced during growth the defect density will become unacceptable and cracking may even occur during cooling. Furthermore, annealing at a temperature of 850 $^\circ$C is required to optimise the optical properties of the devices and homogenise the distribution of dopants and impurities [22], as is the case with other dilute nitrides [82]. Thus, the defect free growth of homogenous material epitaxially on silicon required for devices to achieve and maintain an acceptable level of efficiency requires the precise management of the induced strain and lattice constant mismatch as a function of temperature [28].
Figure 5.6: Lattice constant vs temperature for GaP, Si and BGaP (for boron content of 1 - 4%) between ambient temperature and 1000 K (taken from [18]).

It has been found that small concentrations of boron can be used to tune the lattice constant of GaP to match that of silicon at a chosen temperature [19] as depicted in figure 5.6. However, boron containing III-V materials are poorly understood due to a lack of previous studies and as such the impact on band structure of the inclusion of boron was unknown, making designing a structure incorporating boron containing alloys challenging. Thus N. Hossain carried out a study to determine the the band structure properties of BGaP with a focus on suitability as a cladding or barrier material within GaNAsP laser devices.

It was found through photo-voltage spectroscopy that the direct and indirect bandgaps of GaP decreased from 2.74 eV and 2.2 eV to 2.4 eV and 1.8 eV respectively at room temperature with a boron content of 6%. These bandgaps were determined to be greater than that of GaNAsP to a suitable degree that BGaP could be employed as either cladding or barriers as well as for strain compensation layers. X-ray diffraction and atomic-force microscopy both confirmed the crystalline quality of GaNAsP/BGaP/Si MQW heterostructures, indicating that this material met the requirements of lattice matching to silicon and compensating induced strain. Further, photoluminescence measurements of heterostructures made using these barriers indicated reasonable optical properties, suggesting that the devices show potential for electrical injection lasing. It was also noted that carrier leakage was an issue for devices with GaP barriers and since the BGaP has a slightly lower bandgap the carrier leakage would likely become worse for devices grown on a silicon substrate.

Figure 4.25 presents structure #16195, on which those investigated here are based as well as the band alignments. The direct band gap of Ga(NAsP), conduction band offset and valance band alignment were measured by Sven Liebich et. al. at Philipps-University, Germany. The indirect X, L and spin-orbit spin-orbit (SO) band is calculated using the Vegards law. The Γ, X, L and SO bands of GaAs and GaP required for these calculations are taken from Vurgaftman [102] (values of 2.8 eV (1.51 eV), 2.3 eV (1.98 eV), 2.7 eV
(1.8 eV) and 0.08 eV (0.341 eV) for GaAs (GaP)). The E_ level of Ga(NAsP) alloy is taken from previous investigation of structure #16195 by N. Hossain [66].

<table>
<thead>
<tr>
<th>Material</th>
<th>Thickness (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_{0.033}GaAsP:Zn</td>
<td>100</td>
</tr>
<tr>
<td>B_{0.033}GaP:Zn</td>
<td>600</td>
</tr>
<tr>
<td>B_{0.033}GaP:Zn</td>
<td>600</td>
</tr>
<tr>
<td>GaP</td>
<td>10</td>
</tr>
<tr>
<td>B_{0.06}GaAs_{0.11}P</td>
<td>150</td>
</tr>
<tr>
<td>GaP</td>
<td>3</td>
</tr>
<tr>
<td>GaN_{0.07}As_{0.8}P</td>
<td>5</td>
</tr>
<tr>
<td>GaP</td>
<td>3</td>
</tr>
<tr>
<td>B_{0.06}GaAs_{0.11}P</td>
<td>150</td>
</tr>
<tr>
<td>B_{0.033}GaP:Te</td>
<td>500</td>
</tr>
<tr>
<td>B_{0.033}GaP:Te</td>
<td>1000</td>
</tr>
<tr>
<td>GaP</td>
<td>10</td>
</tr>
<tr>
<td>GaP</td>
<td>60</td>
</tr>
<tr>
<td>Si(001) ex.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.7: Left: The composition of device structure #16195. Right: The band alignments for #16195 at 120 K, 0 energy is taken to be the BGaAsP SO band edge.

Of the devices grown on silicon it was found that structure #16195 (See figure 5.7) was the best performing to date, lasing up to a maximum temperature of 165 K with a $J_{th}$ of 1.6 $kAcm^{-2}$ and a lasing wavelength of 861 nm. In comparison the GaNAsP/GaP devices were found to have $J_{th}$ and lasing wavelength values of 0.54 $kAcm^{-2}$ and 923 nm respectively at the same temperature (165 K). Device performance on a Si substrate is clearly significantly worse, especially considering the devices on a GaP substrate operated up to room temperature.

The characteristic temperature, $T_0$ was found to be unusually large at a temperature of 100 K, with a value of 198 K, similar to the case of the devices on a GaP substrate, again indicative of inhomogeneity within the active region (see the discussion of figure 5.11 for details). $T_1$ was also found to be low at 99 K, below a temperature of 100 K and dropping rapidly to 35 K above this temperature. This, again as was the case with the GaP substrate devices, where $T_1$ was found to be 118 K, dropping to 87 K over the same temperature range indicates the presence of optical losses or a drop in internal quantum efficiency. Superlinear integrated spontaneous emission at 40 K suggests the presence of defects in the GaNAsP/Si devices, which was not observed in the case of the GaNAsP/GaP devices indicating that an increased defect density related to the transferal onto silicon could be partly responsible for the lower performance.

$Z$ analysis gave values of 1.7 and 2.3 at 40 K and 165 K respectively indicating the presence of defects, consistent with other dilute nitride lasers [125]. The GaNAsP/GaP devices were not found to have $Z < 2$
at any temperature, indicating again that the defect density has increased as a result of epitaxial growth on a silicon substrate. The strong increase in Z value with temperature was taken as further evidence of carrier leakage since the current path exhibits an exponential dependence on temperature (see section 2.3.3). It was estimated from the temperature dependent electroluminescence that the radiative, leakage and defect components of the threshold current were ~17%, 43% and 40% respectively at 165 K, indicating that even at cryogenic temperatures, non-radiative recombination dominates these devices.

Pressure dependent electroluminescence again indicated carrier leakage to be a dominant leakage mechanism. However a poor fit to experimental data was found if leakage though the X-minima of the BGaAsP and BGaP barrier and cladding layers was assumed to dominate. Using the measured pressure coefficients it was determined that up 3.8% of nitrogen may have been incorporated into the GaP spacer layers or BGaAsP barriers, a feature again shared with the GaNAsP/GaP devices. Particularly troubling is the potential for strong nitrogen-boron bonds should nitrogen diffuse all the way into the BGaAsP layers, as these have been suggested to act as strong non-radiative centres [82, 16]. Further details of this work can be found in the thesis of N. Hossain, reference [66].

Thus the work of N. Hossain comes to similar conclusions to that of J. Chamings. The main cause of non-radiative recombination for these devices, regardless of substrate is thought to be related to the diffusion of nitrogen from the active region into the surrounding layers. Inclusion into GaP induces an $E_{\text{sub}}$ sub-band in accordance with the BAC model, which may act as a carrier leakage path, leading to high temperature sensitivity and threshold current density. Furthermore, in the case of the GaNAsP/Si devices diffusion of nitrogen through the GaP spacers into the BGaAsP barriers may induce strong boron-nitrogen bonds, acting as localised defect states. In both cases optical losses were determined to be significant.

5.2 The dependence of efficiency limiting processes on GaNAsP/Si quantum well number

Section 5.2 investigates the effect on the performance of the GaNAsP/Si quantum well laser on the number of quantum wells included in the active region and was carried out in collaboration with N. Hossain.

Multiple quantum well laser structures (as opposed to single quantum well in the case of structure #16195 and earlier) are generally used in commercial laser devices due to a number of beneficial factors. As a result, a final design for the monolithic GaNAsP/GaP/Si lasers will include multiple quantum wells. It is therefore of interest to study the effects of additional quantum wells on device performance.

The simple fact that the volume of the waveguide composed of QW material is larger, whilst maintaining quantum confinement causes the optical confinement factor (Γ) to increase (details of optical confinement can be found in section 2.6.3). Carrier confinement may also be increased due to reduced band filling, since the overall density of states is effectively increased proportionately to quantum well number, thus reducing filling for a given injection current. This in turn aids in minimising carrier leakage, a loss mechanism that, as discussed above, has been found in the previous studies of GaNAsP based lasers to be highly prominent. Stronger confinement of carriers as well as an increased density of states at the $E_1$ energy level from which
stimulated emission takes place will in turn lead to an increased differential gain (since gain is a logarithmic function of carrier density in the case of a quantum well), a primary indicator of an efficient laser.

The net effect of all of these properties is an all important reduction in threshold current per quantum well and maximisation of external quantum efficiency (see equations 2.34 and 2.36). The returns of adding additional quantum wells diminishes due to an increase in total threshold current and hole injection issues for a structure of increasing thickness and so in practice devices with between 3 and 9 quantum wells are most common.

The multiple quantum well structure is most simply achieved by repeating the pattern of the single quantum well active region layers up to the required number of wells, as is the case here, one additional factor to take into account however is the separation of the quantum wells. With too little separation the wavefunctions of the carriers may overlap, creating a superlattice structure with sub bands forming as a result of the additional periodicity of the structure, analogously to the formation of the bulk bands within a periodic crystal lattice. With too great a separation however, carrier injection into subsequent quantum wells can be detrimentally affected, thus a material system specific compromise must be reached.

In this section two device structures are investigated. #16052 is almost identical to #16195 but with the addition of a graded composition section to the BGaAsP barriers and a 200 nm thicker n-doped BGaP layer, the band alignment is therefore represented well by the right hand part of figure 5.7. #16053, which is a triple quantum well version of #16052 is shown in terms of tabulated composition (left) and band alignment (right) in figure 5.8. These structures were produced in subsequent growth runs under the same conditions, using the same substrate batches and precursor sources. Thus, any differences in device performance and recombination mechanisms can be attributed to structural differences as opposed to dissimilar growth conditions to a good degree of confidence.
5.2.1 Temperature dependent electroluminescence

Temperature dependent electroluminescence was carried out under pulsed operation with a frequency of 10 kHz and pulse width of 500 ns, giving a duty cycle of 0.5% in order to minimise Ohmic heating to a negligible level. The devices were tested in turn using the closed cycle cryostat experimental arrangement described in section 3.3. As can be seen from figure 5.9 below the maximum operating temperature achieved from both structures was 120 K, the threshold current density per quantum well (using a device area of $1000 \mu m \times 100 \mu m = 0.001 cm^{-2}$) is found to be 0.62 kAcm$^{-2}$ and 0.57 kAcm$^{-2}$ at 40 K, rising to 1.92 kAcm$^{-2}$ and 1.51 kAcm$^{-2}$ for structures #16052 and #16053 respectively. The variation in $J_{th}$ per QW across multiple devices was determined to be $\sim \pm 0.05 kAcm^{-2}$ meaning that the difference at 40 K may not be considered statistically significant, unlike the substantial difference at 120 K. If carrier leakage is assumed to be a dominant loss mechanism in these devices, as it was for previous GaNAsP based lasers then a statistically identical $J_{th}$ per QW at 40 K would make sense, since at this temperature carrier leakage should be comparatively negligible in comparison to SRH and radiative recombination. Whereas, at 120 K
the benefits of the MQW structure in reducing leakage become more relevant, leading to a lower $J_{th}$ per QW. Analysis and quantification of potential loss mechanisms and particularly carrier leakage is carried out in subsequent sections.

It is of interest to note that the $J_{th}$ values are far greater than those of structure #16195, which is almost identical to #16052, yet exhibits a threshold current density of $0.86 \text{kAcm}^{-2}$ (at 120 K) and lased up to a maximum temperature of 165 K. Further discussion of the differences in performance of #16195 and #16052 and #16053 will be presented later in this section.

A small decrease in emission wavelength of $\sim 20 \text{ nm}$ from #16052 to #16053 does suggest that band filling and therefore Fermi-level splitting is reduced in the MQW device, again as predicted, which would be expected to correspond to a lower threshold current density and therefore $J_{th}$. Since carrier leakage has been identified as a major loss mechanism during the investigation of all devices based on the GaNAsP/GaP(/Si) material system, even at cryogenic temperatures, a small decrease in band filling, primarily in the conduction band due to the lower effective mass of electrons compared to holes, would be expected to cause a significant drop in $J_{th}$ as is observed.

Further analysis of the LIs allows the characteristic temperatures, $T_0$ and $T_1$ to be derived from the threshold currents and external efficiencies according to the formulae found in section 2.10, these are presented in figure 5.10 below. The data as a whole exhibits a discontinuity between 85 K and 90 K that sharply reduces $T_0$ and $T_1$ for both structures. In addition the values for each device are broadly similar, with those of #16053 marginally improved. These two points indicate that the dominant efficiency limiting process is common.

Figure 5.9: L-I curves for single (#16052, blue) and multi (#16053, black) quantum well structures as a function of temperature from 40 K - 120 K. The LIs have been plotted with their threshold current per quantum well for ease of comparison between devices.

Further analysis of the LIs allows the characteristic temperatures, $T_0$ and $T_1$ to be derived from the threshold currents and external efficiencies according to the formulae found in section 2.10, these are presented in figure 5.10 below. The data as a whole exhibits a discontinuity between 85 K and 90 K that sharply reduces $T_0$ and $T_1$ for both structures. In addition the values for each device are broadly similar, with those of #16053 marginally improved. These two points indicate that the dominant efficiency limiting process is common.
to both structures, reducing performance, although the advantages of the MQW structure mentioned above lead to slight improvements, but do not address the main source of loss. The discontinuity in both $T_0$ and $T_1$ is indicative of a thermally activated loss process, carrier leakage could cause such a drop in performance due to its temperature sensitivity (see section 2.3.3) and has been previously shown to be a dominant source of loss (see section 5.1.3). Auger recombination, although highly temperature dependent (see section 2.3.2) has a small cross section at both this emission energy and low temperature, allowing it to be discounted.

The low value of $T_1$ (80±4 K and 62±4 K) below 80 K and lower still above 90 K (23±4 K and 16±4 K) could be indicative of optical losses within both devices. However, equations 2.34 and 2.36 also show that a reduction in internal quantum efficiency could be the cause as the stimulated emission current becomes fractionally smaller with increasing temperature. Based on previous findings of carrier leakage limiting GaNAsP based lasers a drop in internal quantum efficiency is the more likely cause as little further indication of optical losses is observed in this study.

Above 90 K the $T_0$ values for devices #16052 and #16053 are 48±5 K and 51±5 K respectively (statistically identical), much less than the operating temperatures. This indicates that radiative recombination is not dominant in these devices, however the values do not align well with the ideal values for any one non-radiative process and so which are dominant is not clear. Below 80 K the $T_0$ value of both devices is significantly greater than the device temperature, for a radiatively, defect or carrier leakage dominated device we would expect $T_0$ values of $T$, $\frac{2}{3}T$ and $\sim \frac{T}{e^\frac{a}{k_B T}}$ respectively. Thus the main recombination processes that have been found to take place in the GaNAsP/GaP/Si quantum well lasers cannot account for these values of $T_0$ alone.
Figure 5.10: Characteristic temperatures, $T_0$ (black) and $T_1$ (blue) for devices of structure #16052 (SQW, solid) and #16053 (MQW, hollow) over the temperature range 40 - 120 K.

A potential explanation for the unusually large $T_0$ (compared to the measurement temperature) values would be an inhomogeneous active region. If the bandgap varies spatially, as depicted in figure 5.11, then at threshold some carriers will be within energy states which meet the Bernard-Duraffourg condition, whilst others will be in higher energy states and will not be able to contribute to lasing. At low temperatures, as is the case here, the carriers in higher energy states may become localised since they lack the thermal energy required to escape these levels and reach those from which lasing is taking place. These carriers may still recombine radiatively but do not contribute to useful stimulated emission and therefore limit device efficiency.

However, as the temperature is increased the carriers localised within the higher energy states may begin to be thermally excited out of these levels and into those from which gain is taking place. Thus, the injection current needs to be increased less than it otherwise would have been were the band gap homogenous, since the previously localised carriers are now taking part in lasing, leading to an increased $T_0$ value. This same process is well known to take place in quantum dots as a result of growth tending to yield a range of dot radii, which behave analogously to an inhomogeneous active region and are known to exhibit similarly large $T_0$ values.

Investigation by both J. Chamings and N. Hossain have presented evidence of the potential diffusion of nitrogen out of the active region[66, 34]. Since the devices investigated here were grown under similar conditions to structure #16195 it is likely that a similar level of nitrogen diffusion is also taking place. If this were the case then the local nitrogen concentration would vary across the quantum well and since the $E_-$ sub band is so sensitive to nitrogen concentration (see section 5.1.2) bandgap inhomogeneity is plausible.
This result would imply that nitrogen diffusion from the active region occurs in both #16052 and #16053, as was the case in previous devices (see section 5.1.3) and that the associated efficiency limiting processes resulting from nitrogen diffusion are likely taking place also. This result alone is not sufficient evidence to draw such a conclusion and so nitrogen diffusion and the associated inhomogeneity of the quantum well band gap are discussed further with reference to subsequent experimental results presented below.

Figure 5.11: Spatial variations in the band gap resulting from proposed local variations in nitrogen concentration, taken from [58].

A Z analysis was carried out as a function of temperature according to the method outlined in section 2.9. The results for both structures are depicted in figure 5.12, the black data representing #16052 (SQW) and the blue data #16053 (MQW). The Z analysis gives an indication of the average carrier density power law dependence of the net recombination taking place as a result of all mechanisms present. It can therefore give further indications of which mechanisms are dominant.

As was the case with the characteristic temperatures the Z values are broadly similar between the two devices, indicating once more that the main efficiency limiting processes are common to both structures and are not strongly effected by the benefits of the additional quantum wells. At 40 K the Z values at threshold, $Z_{th}$, are found to be 1.53±0.07 and 1.54±0.07 (statistically equal) for device structures #16052 and #16053 respectively, indicating that defect related recombination is significant and accounts for a large proportion the threshold current at low temperatures. By 120 K $Z_{th}$ is found to have risen to 2.2±0.07 and 2±0.07 for device structures #16052 and #16053 respectively. The Z value reaches 2 (the ideal value for a radiatively dominated device) and does not appear to be leveling-out, but in fact shows an increasing gradient, indicating the presence of a highly temperature dependent recombination mechanism. As Auger recombination has already been discounted due to both the low temperature and short emission wavelength this leaves carrier leakage as the remaining candidate.

An exponential function was found to fit both data sets well, supporting the hypothesis of carrier leakage being a dominant non-radiative recombination mechanism based on the following analysis. If equation 2.19 is considered it can be seen that the reduction in $E_B$ exponentially increases the leakage current. Increasing carrier density leads to band filling which lowers $E_B$, allowing the substitution $-E_B = \frac{n}{n_0}$ to be made, where
\( n_0 \) is a constant. This gives an equation of the form \( I_{\text{leak}} \approx I_0 \exp \left( \frac{n}{n_0 k_B T} \right) \), which when expanding the exponential gives an equation of the form \( I_{\text{leak}} \approx I_0 (An + Bn^2 + Cn^3 + Dn^4 + En^5...) \).

Although the errors on the absolute values are large enough to make direct comparisons of \( Z \) values statistically unsound the overall trend of a lesser rate of increase of \( Z \) for structure #16053 may be significant. This would be consistent with a reduction in carrier leakage as a result of the increase in quantum well number. The lower \( J_{th} \) and lasing wavelength both indicate a reduced band filling and so it is possible that this would be at least partly responsible for a reduction in carrier leakage through an increase in \( E_B \).

![Figure 5.12: The “Z” parameter at threshold plotted as a function of temperature over the range 40 K - 120 K for structures #16052 (SQW, black) and #16053 (MQW, blue). An exponential function was found to fit well with both data sets indicating the presence of carrier leakage.](image)

Isolating the contributions from each recombination mechanism allows quantification of each, allowing a more precise description of the physical limitations of the devices to be given. Extrapolation from the low injection region of the \( Z \) analysis plots, where \( Z \approx 1 \) to threshold allows an approximation of the defect current contribution at threshold to be made, details of this process can be found in section 2.9.1. The radiative component of the threshold current can also be approximated using the pure spontaneous emission collected from the device substrate using the method described in section 3.2.2. The integrated spontaneous emission at threshold is proportional to the radiative threshold current density (\( J_{\text{rad}} \)). The integrated spontaneous emission at each temperature is measured at threshold and normalised to \( J_{th} - J_{def} \) at 40 K in order to give \( J_{\text{rad}} \) as a function of temperature. At 40 K, \( J_{th} - J_{def} \approx J_{\text{rad}} \) since at low temperature we assume carrier leakage (and Auger recombination) are negligible due to a strong temperature dependence. Thus the value of \( J_{\text{rad}} \) calculated using this method is an approximation of the maximum radiative component of the threshold
current density at each temperature.

The radiative component of the threshold current density, \( J_{\text{rad}} \) and the threshold current density with the defect component removed, \( J_{\text{th}} - J_{\text{def}} \) are plotted together as a function of temperature for both structures in figure 5.13 and normalised at 40 K. An exponential function is found to fit well to the \( J_{\text{th}} - J_{\text{def}} \) data, further supporting the previous conclusions of carrier leakage being one of the two main sources of loss in these devices along with SRH recombination. Plotting these two parameters together gives an indication of at what temperature carrier leakage becomes a significant source of loss. Below 50 K the gradients of the two data sets are approximately equal, suggesting that below this temperature carrier leakage is minimal, as was predicted when determining \( J_{\text{rad}} \) based on the strong temperature dependence of carrier leakage. When the temperature reaches 60 K the deviation in gradients is still small but can be seen to increasing rapidly. Beyond 60 K the increase in gradient is strong and appears to be approximately exponential in form indicating that at temperatures above 60 K a strongly temperature dependent recombination mechanism other than radiative recombination is taking place that is most likely carrier leakage (since Auger recombination is expected to be negligible). That carrier leakage would start to become significant at such low temperatures is indicative of the carrier confining energy barrier height being low and consequently that the leakage level is close to the quantum well \( E_1 \) level in the \( E_- \) sub band.

Figure 5.13: Extrapolation of the low injection regime of the Z analysis plots (see section 2.9.1) where \( Z = 1 \) gives the approximate threshold defect current density fraction. Deducting this from the threshold current density gives the net contribution of the radiative and carrier leakage currents at threshold (solid symbols). Normalising the integrated spontaneous emission at 40 K (where carrier leakage can be assumed to be negligible) to \( (J_{\text{th}} - J_{\text{def}}) \) yields the approximate radiative components of the threshold current density (hollow symbols).
Since Auger recombination is thought to be minimal as well as multiple sources of evidence for carrier leakage, the normalised leakage component of $J_{th}$ is calculated as: $J_{leak} = J_{th} - J_{def} - J_{rad}$. Combining the ratios of the normalised values with the measured $J_{th}$ values yields the $J_{th}$ components. The threshold components for both structures #16052 and 16053 are plotted as a function of temperature in figure 5.15, along with those of structure #16195 (taken from [66]) for comparative purposes.

The defect current is increasing with temperature, which is indicative of the defects being shallow. The nitrogen atoms within the active region are classified as deep centres and so would not be considered to be the shallow defects implied here. However, if the hypothesised nitrogen diffusion is taking place then the N concentration and therefore $E_\text{L}^-$ energy will vary across the well. A lower N concentration will cause states of a higher energy to form which do not contribute to lasing and thus may act as a type of defect state. States may also exist at the GaNAsP/GaP interface in the form of GaNP $E_-^\text{L}$ levels with which thermally excited carriers could interact, also giving the form of a thermally activated defect. Thus carriers may increasingly combine with these additional defect states as their thermal energy increases. This proposed mechanism is depicted approximately in figure 5.14.

Figure 5.14: The defect states formed by the proposed nitrogen diffusion from the QW(s). The normal An type SRH recombination in the QW(s) and thermally activated “leakage like” recombination are shown. Boxed is the corresponding assumed nitrogen concentration distribution.

The implication here would be that additional thermally activated defects are formed near the GaNAsP/GaP interface in addition to the “normal” defects that would generally be expected to cause SRH recombination. SRH recombination from these states, being dependent on the carrier density with sufficient energy to occupy...
them would then be expected to have a form somewhat like a carrier leakage type process. The nitrogen that has diffused from the QW(s) would create states in the GaP or even BGaAsP layers depending on how far it diffused that could potentially act as a carrier leakage path, this is also depicted in figure 5.14. The approximate separation of the lowest barrier states from the QW ground state (\(\sim 33\) meV), corresponding to a N concentration of \(\sim 5\%\) at the GaNAsP/GaP interface is approximated below with reference to figure 5.17.

The reduction of both SRH recombination and carrier leakage current with the addition of multiple quantum wells likely stems from the increased optical confinement factor and differential gain yielding a lower threshold current density, on which carrier leakage is strongly dependent. Comparing structures #16052 and #16053 with #16195, previously investigated by N. Hossain it is clear that both leakage and defect components of the threshold current are much larger whilst the radiative component is almost identical. Thus there must be some differences between these structures that have not yet been accounted for, since they should be structurally almost identical and grown under similar conditions. In the case of #16195 the defect, radiative and leakage fractions were demonstrated to be 52\%, 31\% and 17\% respectively at 120 K[66].

![Figure 5.15: The defect, radiative and carrier leakage components of the threshold current density plotted as a function of temperature for structures #16052, #16053 as well as #16195 for comparative purposes. Based on multiple sources of evidence for the presence of carrier leakage the leakage component is determined as: \(J_{\text{leak}} = J_{\text{th}} - J_{\text{def}} - J_{\text{rad}}\). Data for #16195 taken from[66].](image)

Replotting the threshold components as a percentage of the total threshold current density allows a better representation of how the proportions of current associated with each mechanism vary with temperature. From figure 5.16 it can be seen that across the entire temperature range SRH recombination is dominant, accounting for \(>50\%\) of injected current all cases, and only dropping slowly. Leakage on the other hand
accounts for a negligible proportion at low temperatures as is expected but rises rapidly by 120 K. It would be expected to become the dominant loss, overtaking the SRH component well before room temperature, assuming the devices operated at such temperatures. It is therefore clear that although SRH recombination is a major source of loss it is carrier leakage that limits the maximum operating temperature of all three structures. It also appears as though the source of SRH recombination and carrier leakage is the same in all three structures as they follow similar trends. The values are 120 K for all three structures as well as their approximate errors are plotted in table 4 for ease of comparison.

Comparing #16052 (SQW, blue) to #16053 (MQW, black) the radiative current fraction is considerably lower, as one would expect based on the aforementioned confinement factor and band filling improvements. The defect and carrier leakage current fractions however do not differ by a statistically relevant magnitude as a result of the reasonably large error associated with their determination. Combining these two facts the conclusion is that the additional QWs benefit the device but it is not clear whether this is a result of a reduction in SRH recombination, carrier leakage or a combination of both mechanisms and to what degree.

The large $J_{def}$ fractions in both cases even with significant errors fit the the low $Z_{th}$ values (<2 for most temperatures) presented in figure 5.12, which indicated a significant threshold current fraction associated with SRH recombination. The lower SRH and carrier leakage fractions and slower growth of carrier leakage in the case of #16195 reaffirms the previous conclusion of the greater performance of #16195 being the result of reduced non-radiative recombination. This information however again does not indicate as to why, and is the subject of further investigation later in this section.

<table>
<thead>
<tr>
<th>Structure ref #</th>
<th>no. QWs</th>
<th>$J_{def}$ (%)</th>
<th>$J_{rad}$ (%)</th>
<th>$J_{Leak}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16052</td>
<td>1</td>
<td>65 ± 4</td>
<td>8 ± 2</td>
<td>27 ± 5</td>
</tr>
<tr>
<td>16053</td>
<td>3</td>
<td>58 ± 4</td>
<td>18 ± 2</td>
<td>24 ± 5</td>
</tr>
<tr>
<td>16195</td>
<td>1</td>
<td>52 ± 4</td>
<td>31 ± 3</td>
<td>17 ± 5</td>
</tr>
</tbody>
</table>

Table 5.1: The approximate threshold component fractions at 120 K for structures #16052, #16053 and #16195. Data for #16195 taken from[66].
Figure 5.16: The defect, radiative and leakage components expressed as a percentage of the total threshold current density. Blue, black and red data represent structures #16052, #16053 and #16195 respectively with square, circular and triangular symbols representing defect, radiative and leakage components respectively, as was the case in figure 5.13. Data for #16195 taken from [66].

Using the defect currents as a function of temperature from figure 5.15 a plot of $\ln(J_{\text{leak}})$ vs $1/T$ can be produced, the slope of which is the carrier leakage activation energy over Boltzmann’s constant in accordance with equation 2.19. This is shown in figure 5.17 for all three structures, with the activation energy found to be approximately 33 meV in all three. This information was added to figure 5.14 to give a representation of where the leakage levels within the GaP layers would lie. Using this energy along with equation 5.1 for the $E_-$ state in GaNP gives a nitrogen concentration of $\sim 5 \pm 0.4\%$ at the GaNAsP/GaP interface, which would fall with distance into the GaP potentially BGaAsP layers. How much of a loss of nitrogen from the active region this would represent depends on how much further the nitrogen diffused and at what rate the concentration falls off. The degree of nitrogen diffusion is discussed further with reference to the hydrostatic pressure dependence investigation of section 5.2.2.
5.2.2 Hydrostatic pressure dependent electroluminescence

Temperature dependence analysis has indicated that both SRH recombination and carrier leakage are dominant loss mechanisms in these devices as well as allowing the degree to which each of these processes take place to be quantified. The hypothesis of both SRH recombination and carrier leakage being linked to nitrogen diffusion from the quantum wells, as was concluded in previous studies by N. Hossain and J. Chamings has been presented based on this analysis but as yet requires further evidence to sufficiently substantiate.

The effect of applying high hydrostatic pressure is presented here, it allows reversible changes to the band alignments to be made, giving a complimentary method of investigating non-radiative recombination mechanisms (see sections 2.11, 3.3 and 3.4 for details of this experimental method). As this method is based on changes to band alignments, figures 5.7, 5.8 and 5.14 are useful references to the ambient pressure situation and will be used to help investigate potential carrier leakage paths.

The lasing line peak wavelength was observed as a function of pressure in order to track the change in quantum well band gap and to determine its pressure coefficient. This is plotted in figure 5.18 for temperatures of 80 K and 120 K along side the nitrogen defect levels ($E_N$) and GaAsP Γ-point band gaps ($E_G(GaAsP)$) at both temperatures for comparative purposes. It can be seen that the pressure coefficients for structures #16052 and #16053 are $61\pm2 \text{ meV GPa}^{-1}$ and $39\pm2 \text{ meV GPa}^{-1}$ respectively at both temperatures. These can be used in conjunction with the pressure coefficients of the barrier band gaps in order to approximate the activation energy, $\Delta E_b$ of the leakage path in accordance with equation 2.4.3 from section 2.11.2. In both the
barrier and strain compensation layers the minimum band gap occurs at the X-minima and so the pressure coefficient of these is used in the following calculations. The pressure coefficients of \((B_{0.03}Ga)(As_{0.11}P)\) and \((B_{0.03}Ga)P\) are calculated using interpolation according to Vegard’s law to be \(-14.8\) meV\(\text{GPa}^{-1}\) and \(-14.7\) meV\(\text{GPa}^{-1}\) respectively, nearly identical.

Thus the carrier leakage path into the X-minima would be expected to exhibit an activation energy of \(\sim 71\) meV\(\text{GPa}^{-1}\), this value can be compared to that determined experimentally and used to approximate the degree of leakage through this path that would be expected to take place. The difference in pressure coefficients between the two structures is minimal, as would be expected since they should have identical quantum wells, only differing in number. The offset in lasing energy is accounted for by the proposed reduction in band filling and electron quasi-Fermi-energy discussed at the beginning of this section. The reduction in lasing energy with temperature reflects the thermal expansion of the lattice. A slight super-linearity is observed in the experimental data, this is thought to be a result of the increasing \(J_{th}\) with pressure (see figure 5.19) causing a blue shift in the emission (measured at \(1.1 \times J_{th}\) for each pressure).

Figure 5.18: Lasing energy plotted as a function of hydrostatic pressure for structures \(\#16052\) and \(\#16053\) as well as the calculated nitrogen levels (\(E_N\)) (using the BAC model) and GaAsP gamma band gaps (\(E_G(GaAsP)\)) at 80 K and 120 K. Note the break in the y-axis separating the experimental data and theoretical values.

The threshold current density, \(J_{th}\) (solid symbols) as well as the defect subtracted current density, \(J_{th} - J_{Def}\) (hollow symbols) are plotted in figure 5.19. In order to calculate \(J_{th} - J_{Def}\) the same method was used as per figure 5.13. Also included is the theoretical maximum radiative current, \(J_{rad}\) (red line) normalised to atmospheric pressure, and follows the trend, \(J_{rad} \propto E_g^2\), with the pressure dependence of \(E_g\) determined

137
from figure 5.18 above. Finally, the theoretical threshold current density assuming 100% leakage is plotted in green. This was determined assuming that the leakage path is between the lasing level of the quantum well and the X-minimum of the barriers, with an activation energy of $\sim 71 \text{ meV} \cdot \text{GPa}^{-1}$ as calculated above. This activation energy is then substituted into equation 2.4.3 in section 2.11.2 in order to calculate the hypothetical 100% leakage current density as a function of pressure, normalised to atmospheric pressure.

The three main current paths thought to be present in these devices are SRH, radiative and carrier leakage. The dependence on pressure of pure radiative recombination and leakage into the X-minima are plotted and SRH recombination, being almost entirely pressure independent would be represented by a horizontal line along the x-axis. Both sets of data deviate from ideal radiatively dominated behaviour, with the threshold current density of #16052 increasing by 11% and 25% at 80 K and 120 K respectively and those of #16053 by 8% and 17%.

This result alone does not necessarily indicate increased radiative recombination or reduced carrier leakage in #16053 since increased defect recombination would reduce the rate of increase of $J_{th}$ with pressure. The defect subtracted data (assuming all defects are within the quantum well(s)) therefore is useful for estimating which processes (and to what degree) have been affected by the use of the MQW active region. From the defect subtracted thresholds we can see that all data sets exhibit an increase in gradient as is to be expected after removal of the pressure independent component.

Further to this it is also clear that the relative difference between the gradients of the data sets from each structure have remained similar indicating that the greater gradient of the data from #16052 relates to an increased carrier leakage current fraction and reduced radiative current fraction compared to #16053. This finding is in good agreement with the results of the temperature dependent electroluminescence results presented above which find #16053 to have reduced non-radiative recombination in comparison to #16052.
With a known defect component it is possible to fit the normalised experimentally determined threshold current densities by interpolation of the radiative, defect and carrier leakage components. Since the defect component is (approximately) known this is held constant and the carrier leakage and radiative recombination components are varied to achieve the best fit.

The result of such a fitting for structure #16052 is a leakage component of approximately 3\% and 9\% at 80 K and 120 K respectively and approximately 2.5\% and 7\% in the case of structure #16053. Comparing these to the values determined from the temperature dependence study of 27\% and 24\% for #16052 and #16053 at 120 K respectively, there is a considerable discrepancy. The poor temperature dependent performance of these devices and temperature dependent analysis indicate strongly that carrier leakage is significant in these devices and so the small values calculated here are indicative of leakage into the X-minima contributing minimally to net carrier leakage, if at all. The finding of negligible leakage into the X-minima is consistent with studies of structure #16195 and previous investigations on both Si and GaP substrates[34, 66].

With the quantum well to barrier X-minima carrier leakage path not able to account for the observed leakage it is necessary to approximate the carrier leakage activation energy of the as yet unidentified leakage path. This can be achieved through fitting of the data once more, this time with the use of equations 2.43 and 2.44 in section 2.11.2. By doing so leakage activation energies for structure #16052 of approximately -29 meV/\text{GPa}^{-1} and -28 meV/\text{GPa}^{-1} at 80 K and 120 K respectively are calculated. In the case of structure #16053 the values are approximately -41 meV/\text{GPa}^{-1} and -39 meV/\text{GPa}^{-1}.

Combining these with the pressure coefficients of the quantum wells gives approximate leakage level
pressure coefficients of $58 \text{meV} \text{GPa}^{-1}$ and $59 \text{meV} \text{GPa}^{-1}$ at 80 K and 120 K respectively in the case of #16052 and approximately $55 \text{meV} \text{GPa}^{-1}$ at both temperatures in the case of #16053. These pressure coefficients are clearly considerably different that those of the barrier X-minimum path, further supporting the conclusion of this leakage path playing little role in these devices.

The other potential leakage path proposed in this study was one resulting from nitrogen diffusion into the barriers, which would form an $E_-$ sub band in accordance with the BAC theory with an energy considerably lower than that of the X-minima, providing a more accessible carrier leakage path and lower activation energy (see figure 5.14). The nitrogen defect level has been demonstrated to be relatively pressure insensitive\cite{73} as illustrated in figure 5.18. This information along with the known pressure coefficient of GaAsP, also plotted in figure 5.18 can be used to estimate the the fraction of nitrogen required to yield the observed pressure coefficient.

Flanking the GaNAsP quantum well(s) are GaP buffer regions designed to separate diffused nitrogen from boron in the barriers in order to prevent the formation of boron nitride defects. The pressure coefficient of GaNP as a function of nitrogen concentration has been documented in \cite{86} and is used here along with equation 5.1 in section 5.1.2 to determine the nitrogen concentration required within the GaP buffers to yield an $E_-$ level with the calculated leakage activation energy.

Doing so gives a nitrogen concentration of approximately $4.4\pm0.3\%$ and $4.3\pm0.3\%$ for #16052 and #16053 respectively within the GaP buffer layers. These fit reasonably well with the value $5\pm0.4\%$ estimated from the temperature dependence study using figure 5.17. If the nitrogen concentration within the quantum well is calculated using the band anti-crossing model (equation 5.1) then values of $5.9\%$ and $6\%$ are found in the cases of #16052 and #16053 respectively. If 7\% nitrogen was initially incorporated into both structure #16052 and #16053 then it is implied that $\sim1.1\%$ and $\sim1\%$ respectively out of the initial 7\% has diffused out of the active region and into the adjacent layers.

As is shown in figure 5.14 the diffused nitrogen would be expected to have a decaying concentration with distance from the QWs. Thus, although only around 1/7th of the nitrogen in the QW is expected to have diffused out concentrations of $\sim 4 - 5\%$ as calculated could be achieved in a thin layer at the interface, with the value decreasing into the GaP layers. This would still cause a leakage path to form with the observed activation energy, with recombination taking place within the GaP spacers. Based on these observations it seems unlikely that significant concentrations of N would diffuse all the way to the BGaAsP barriers, suggesting that the GaP spacers successfully isolate the boron from nitrogen atoms.

\textbf{5.2.3 Conclusions}

The main conclusion of section 5.2 is that both structures #16052 and #16053 are limited by a combination of strong SRH recombination and increasingly as the temperature rises, carrier leakage (the mechanism which limits the operating temperature to 120 K). $T_0$ values of $48\pm5$ and $51\pm5$ K at 120 K indicate a highly temperature sensitive $J_{th}$, and since Auger recombination should be negligible at an emission energy of $\sim1.5$ eV carrier leakage is the probable cause. The $T_1$ values of $16\pm4$ and $23\pm4$ K at 120 K, suggest a low internal
quantum efficiency since there has been no evidence of optical losses taking place to a considerable degree, again indicating the presence of carrier leakage in the absence of Auger recombination.

The low Z values at 40 K of 1.53$\pm$0.07 and 1.54$\pm$0.07 indicate the presence of a significant proportion of SRH recombination, with approximation of $J_{def}/J_{th}$ from extrapolation of the low injection, $Z \approx 1$ regime giving a values of 65$\pm$4 and 58$\pm$4 %. The Z value rises to 2.2$\pm$0.07 and 2$\pm$0.07 at 120 K, with a slope found to have an exponential form, further supporting the hypothesis of the onset of carrier leakage with temperature. The closest energy level to the QW conduction band ground state is the X-minimum of the GaP spacer layers, making this the most probable carrier leakage path. However, fitting the pressure dependence data gave only 7-9 % leakage fraction, which did not fit the 24 - 27 ($\pm$5) % estimated from the temperature dependence study or the performance of the devices in general. Thus this leakage path is thought to have a minimal contribution to the total carrier leakage at most.

Previous studies of GaNAsP based lasers on both GaP and Si substrates have suggested the diffusion of nitrogen from the QWs in the barriers, leading to the addition of GaP spacers to prevent the formation of BN non-radiative centres at the GaNAsP/BGaAsP interface. The large $T_0$ values of 120$\pm$5 and 132$\pm$5 K at 40 K are indicative of an inhomogeneous active region, which could be caused by the diffusion of nitrogen out of the QW. The temperature dependent increase in $J_{def}$ suggests the presence of shallow defects above the QW ground state, into which carriers are thermally excited to an increasing degree as the temperature is raised. Higher energy states in the vicinity of the GaNAsP/GaP interface could be formed as the GaNAsP $E_-$ level increases and the induced GaNP $E_-$ level decreases in energy as a result of the N diffusion (See figure 5.14).

Approximating the carrier leakage activation energy allows the GaNP $E_-$ level to be calculated using the BAC model. Such approximations based on temperature and pressure dependence studies gave similar N concentrations at the interface of 5$\pm$0.4 and 4.3 - 4.4$\pm$0.3 % respectively, which are reasonably consistent based on the errors of the approximation. Finally, taking the emission energy and using this with the BAC model to calculate the QW N% gives 5.9-6%, 1-1.1% less than the 7% added during growth, suggesting that this N has been diffused into the surrounding layers. Direct evidence of the diffusion of nitrogen has not been found here, it does however fit with several results from this study.

Comparing the single to multi QW devices it has been found that there is a marginal improvement in threshold current (1.92 - 1.51 kAcm$^{-2}$ per QW at 120 K) and efficiency with the use of a multi quantum well structure. The improvement likely stems from the distribution of carriers between multiple wells, minimising the required quasi-Fermi level splitting. Furthermore, increasing the physical size of the active region increases the optical confinement factor, reducing the threshold carrier density required to reach lasing threshold. The lower band filling reduces the effect of highly carrier density dependent recombination mechanisms such as carrier leakage, increasing the fraction of carriers that undergo radiative recombination in the quantum well.

Thus the recommendation based on these conclusions are that firstly the focus of further study to be on the reduction of the defect density within the active region, since defects appear to be the main limiting factor in all GaNAsP based devices. Secondly, that a study be carried out to confirm or dismiss the hypothesis
of nitrogen diffusion and the resultant formation of defect states within the GaP spacers. If this processes is confirmed to be taking place to a significant degree (based on past results it is likely happening to some extent) then growth conditions will need to be optimised with a focus on its reduction. This could potentially eliminate the proposed thermally activated defects and carrier leakage observed in this study and allow operation at greater temperatures.

5.3 The effect of silicon surface orientation on GaNAsP/GaP/Si QW laser non-radiative recombination

This section investigates the effect on the performance of the GaNAsP/Si quantum well laser of the surface orientation of the silicon substrate and was carried out in collaboration with N. Hossain. Both experimental data collection and analysis of said data were carried out in tandem at the time of the investigation.

The conclusions of section 5.2 centre the presence of large defect densities in and around the active region causing the majority of recombination to be non-radiative, in addition to carrier leakage as the operating temperature is increased. Nitrogen diffusion from the quantum well, having been observed in GaNAsP based devices to a great extent previously (see the discussion of figure 5.5) could also be taking place in this generation of devices. It was found to fit with the observations of section 5.2 and is a plausible explanation (although not definitively evidenced) for the observation of additional thermally activated defects and carrier leakage.

It is also known from the studies of N. Hossain that the GaNAsP/GaP devices out perform those grown on a silicon substrate significantly; with increased maximum operating temperature, reduced threshold current density and reduced non-radiative recombination [66]. The issues of strain discussed in sections 1.1 and 5.1.3 if not properly addressed will reduce efficiency through increased non-radiative recombination, primarily through defects. Low temperature Z values of > 2 were also reported by N. Hossain indicating a minimal contribution to non-radiative recombination, compared to as much as 65% of the total recombination for GaNAsP based devices grown on Si. Thus it is reasonable to conclude that integration on Si leads to the formation of a large density of defects that severely limits lasing efficiency.

Therefore optimisation of the growth process on Si is essential in order to achieve a commercially viable electrical injection lasing operating efficiently at room temperature. The effect on non-radiative recombination of varying the III-V/Si interface properties and growth conditions is therefore studied to determine an optimal method for monolithic integration on Si that minimises the resultant generation of defects.

5.3.1 Defect generation at the polar-non-polar GaP/Si interface

The quality of the interface between the III/V semiconductor and underlying silicon substrate heavily impacts whether subsequent epitaxially deposited layers are homogenous and defect free. A key issue when growing GaP on Si is that GaP is polar, which is to say the GaP “molecule” is dipole-like in nature, whilst silicon is non-polar. In a continuous GaP crystal the polar nature of the bonds is not an issue since the Ga-P bonds
are each joined front to back to others leading to macroscopic charge neutrality.

However, since Si is non polar achieving charge neutrality at the III-V/Si interface is not guaranteed. Charge neutrality requires an equal number of Ga-Si and P-Si bonds; if this condition is not met a charge builds up at the interface since group III Ga acts somewhat like an acceptor and group V P somewhat like a donor to group IV Si. The surface charge induces an electric field with a macroscopic range, causing perturbations to the GaP lattice as further layers are grown. This electric field can cause three dimensional island-like structures to form in what should ideally be a “two dimensional” structure; these are depicted in the TEM images displayed in figure 5.20. Coalescence of such islands as growth continues can lead to the generation of stacking faults (5.20(b)) and twins[123]. In addition, the induced charge can also yield oppositely charged defects, which form in order to return the net charge to zero[83].

![Figure 5.20](image)

**Figure 5.20: TEM images of GaP grown on Si under poorly optimised growth conditions, planar defects are marked by arrows. The dark sections of (b) and (c) represent anti phase domains (APDs). (a): Surface view from [001] direction indicating planar defects. (b): Cross sectional view for the case of an excess of Si-P bonds, stacking faults visible within the main phase of the GaP buffer layer. (c): Cross sectional view for the case of an excess of Si-Ga bonds, stacking faults contained within the dark (APDs) of the GaP buffer layer. Taken from [2].**

### 5.3.1.1 Anti phase domains (APDs) and anti phase boundaries (APBs)

The most commonly observed crystal defects found in GaP when grown epitaxially on silicon are anti phase domains (APDs), these form commonly at monatomic discontinuities on the Si surface (see figure 5.21) but may also form on an atomically smooth surface[67]. As depicted in figure 5.21, APDs are regions of a crystal within which the polarity of the bonds is opposite to that of the main phase.

At the interface between APDs and the main phase a so-called anti phase boundary (APB) forms, along the boundary bonds between similar atoms are forced to form since the phases either side of the APB are dissimilar. If the bonding at the APB is stoichiometric (equal numbers of Ga-Ga and P-P bonds in this case) then macroscopically the effect of the similar bonds is balanced. However, locally the fact that the similar bonds are not of equal strength to the Ga-P bonds, causes the lattice in the locality of the APB to be deformed as a local strain is induced[64]. On the other hand if the APB is non-stoichiometric then an imbalance of similar bonds will lead to the formation of charged defects, which may act as scattering centres for charged carriers[62].

143
From figure 5.21 it can be seen that monatomic steps in the Si surface will generate APDs in the epitaxial GaP, the atomic sites of a pseudomorphic GaP lattice will alternate between Ga and P over a monatomic Si step length, reversing the polarity of the bonds either side of the step. Monatomic mesas in the Si surface will also create APDs, however since these form APBs which are not parallel to the [001] growth direction they can close if sufficient thickness of GaP is deposited, after which the lattice will continue to grow according to the main phase. The relaxation of APDs during annealing or overgrowth of subsequent layers tends to form stacking faults, twins and charged defects and so APD density in GaP buffer layers is a strong indicator for the achievable material quality and defect density of a complete device[2].

![Figure 5.21: Monatomic discontinuities in the Si surface yield APDs in epitaxial overgrowth of GaP. Bonds between similar atomic species mark the APBs][2]

Since Ga and P atoms alternate, whilst monatomic steps lead to APDs, double steps do not. Double steps cause the atoms at the discontinuity to align with their correct positions within the lattice and yield a “perfect” lattice structure. As a result, if the Si surface can be grown in such a way as to only exhibit double atomic steps then the density of APDs and by extension defects should be minimised. However, in practice a perfectly double stepped surface is not easily achieved on an exact Si (001) surface since monatomic steps are energetically favourable [70] and therefore the density of APDs can be reduced only so far.

It has however been demonstrated [76] that the favourability of double steps on the Si surface increases as a function of miscut angle from the [001] direction towards the [110] direction. It is therefore possible to produce a surface on which a Si buffer, preferably forming double steps, can be grown. The issue with using a large miscut is that the GaNAsP/GaP/Si lasers are designed to be integrated into Si microelectronic circuits. In order to be suitable for optoelectronic integration purposes, laser growth and processing must be CMOS compatible. Section 1.2. states that CMOS processing methodology stipulates that the [001] surface orientation should be used with a maximum acceptable miscut of ±0.5° [83]. Therefore, to maintain CMOS compatibility the Si substrate used for these devices should also adhere to these limits. To accommodate a
larger miscut the processing time, complexity and therefore cost of the integrated device would be increased, making this option untenable.

5.3.1.2 Optimisation of the GaP buffer for defect free III-V active material growth on Si  Due to its small lattice mismatch to Si, the growth of GaP on Si has been readily achievable and demonstrable for some time under the correct growth conditions [49]. However, the presence of APDs, APBs and the generation of associated defects is problematic for efficient lasing operation from devices based on such layers. A thin layer of GaP can serve as a buffer between the non-polar Si and the polar III/V materials provided the material can be grown defect-free and with minimal APDs and APBs.

If this can be achieved then subsequently grown epitaxial layers, in which carrier injection and photon generation take place will be isolated from the problematic III/V/Si interface and may be grown to a suitable crystalline quality for efficient lasing to take place[2]. Previous studies have reported favourable formation of double steps for a Si surface miscut of 0.1° from the (001) towards the (110) direction and thus reduced the formation of APDs, APBs and defects, however, beyond 0.2° single steps begin to dominate once more[21]. Thus, there is a small window of miscut angle that favours the formation of the double steps required for defect-free growth, whilst conforming to the CMOS processing requirement of miscut angle being no larger than 0.5°.

In this section device performance is investigated for two device structures, A and B, of differing Si surface orientation, both of which meeting the CMOS processing requirement in terms of acceptable miscut. By doing so the optimal surface orientation for device performance can be determined, and by implication from lasing performance, the quality of the GaP buffer as a result of said orientation. Both devices have a 0.1° miscut from the (001) direction towards the (110) direction but device A also has a 30° miscut (rotation) in the [001] plane away from the (110) direction. Aside from the difference in miscut angle the structures are otherwise identical to #16052 and #16195 as detailed earlier in figure 5.7.

5.3.2 Temperature dependent electroluminescence

Temperature and pressure dependent electroluminescence were carried out using the same methodology and experimental arrangement as that of the investigation detailed in section 5.2 and again under pulsed operation with a duty cycle of 0.5% to minimise Ohmic heating. Structure #16052, investigated in section 5.2 is identical to device B other than having been produced as a part of a separate growth and processing batch and device A differs only in the substrate miscut angle. Thus analysis in this section focuses mainly on a comparison of the lasing behaviour and physical properties of devices A and B and is therefore somewhat less detailed than section 5.2 to avoid unnecessary repartition of analysis.

The $J_{th}$ values at 40 K were found to be 1.35 $kAcm^{-2}$ and 0.62 $kAcm^{-2}$ for A and B respectively. The maximum operating temperatures and associated threshold current densities were found to be 3.29 $kAcm^{-2}$ at 80 K and 1.12 $kAcm^{-2}$ at 120 K in the cases of structures A and B respectively. For comparative purposes the threshold current density of device B at 80 K was determined to be 0.873 $kAcm^{-2}$. Since, as mentioned
before individual investigation provides no information not otherwise covered in section 5.2 the investigation is limited to a maximum temperature of 80 K hereafter.

From these two results alone there is a clear indication of the superiority of structure B compared to A. It is expected that this is linked to an improvement in crystalline quality of the GaP buffer, resulting from the favoured formation of double steps on the Si surface, as discussed above. Since the two devices were grown and processed in the same batch it is unlikely that growth conditions or processing differences are the cause of the performance discrepancy, but further investigation below is required to confirm this and quantify the defect recombination fraction.

Plotting the natural logarithm of the threshold current and natural logarithm of the slope efficiency against temperature yields the graph presented in figure 5.2 below. From the inverse gradients the characteristic temperatures $T_0$ and $T_1$ are determined and are found to be $45\pm3$ K and $119\pm3$ K ($T_0$) and $33\pm10$ K and $60\pm10$ K ($T_1$) in the cases of structures A and B respectively. Over the temperature range of 40-80 K the theoretical characteristic temperature values one would expect for defect, radiative and Auger recombination are approximately 27-53 K, 40-80 K and 6-10 K respectively (based on the discussion of equation 2.40 in section 2.11).

The values for device B are consistent with those of #16052 presented in figure 5.10 and discussed in section 5.21 above ($120\pm5$ K ($T_0$) and $62\pm4$ K ($T_1$)) suggesting that growth and processing differences were minimal between these batches and that previous analysis and conclusions should be relevant. Thus the $T_0$ value of $119$ K considerably exceeding the operating temperature remains an implication of an inhomogeneous active region as discussed before, and the fairly low $T_1$ of $60$ K indicative of internal losses and/or strong optical absorption (see section 5.21 for details).

The $T_0$ value for device A of $45$ K varies from slightly larger than the ideal radiative value at 40 K to considerably less at 80 K. Thus, assuming a similar degree of inhomogeneity to device B it is implied that the impact of either defect of Auger recombination must be larger to yield the low value observed here. Since this investigation was carried out at low temperatures as well as taking into account the emission energy of $\sim1.54$ eV, Auger recombination should however be negligible.

Therefore, this result implies increased defect related recombination and in turn an increased defect density within the active region. This is consistent with a less optimised Si surface orientation in the case of device A yielding an increased APD density as detailed in section 5.3.1. The $T_1$ value for device A is also considerably lower than that of device B at 33 K. A lower value is indicative of again even greater optical losses or a lower internal loss, internal losses through scattering increasing as a result of an increased defect density would be consistent with the hypothesis that device A has a less optimised Si surface orientation that device B.
Carrying out the Z analysis for both devices as per section 5.2.1 for structure #16052, and according to the method detailed in section 2.9.1 yields the plot presented in figure 5.23. The threshold Z values are found to vary from 1.32±0.04 to 1.39±0.04 and 1.53±0.04 to 1.65±0.04 over the temperature range of 40 K to 80 K for devices A and B respectively. Once more the values of Z for device B are near identical to those of structure #16052 presented in figure 5.12, indicating that the analysis carried out in section 5.2.1 is again relevant in the discussion of these devices.

We first recall that the expected values of Z for defect and radiative recombination are 1 and 2 and estimate that carrier leakage should be reasonably minimal at such low temperatures. Based on this, and the analysis carried out in section 5.2, that both devices exhibit values of considerably less than 2 is clear indication that defect-related recombination is highly significant in both devices. The lower value of device A compared to B must be as a result of either reduced carrier leakage or increased defect recombination to achieve such a value. Having observed an increased threshold current density for device A and indications of increased defect density from the analysis of figure 5.21 it is clear that the former of the two possible explanations makes little sense and can be discarded. The greater Z value of device B is consistent across the temperature range and makes little difference to the slope of the plot, indicating that the dominant sources of loss remains the same but that SRH recombination has been reduced by optimisation of the Si surface orientation.
Using the same methodology as was applied to produce figure 5.15 in section 5.2.1 the threshold current density was split into its components as a function of temperature for both devices and is presented in figure 5.24 below. The defect component is approximated from extrapolation of the low injection regime of the $Z$ analysis plot where $Z \approx 1$ to threshold and the radiative component from the pure spontaneous emission collected through the substrate. Since Auger recombination is assumed to be minimal in these devices based on the lasing wavelength and previous analysis from section 5.2, the leakage component is then determined to be $J_{\text{leak}} = J_{\text{th}} - J_{\text{def}} - J_{\text{rad}}$.

![Graph of $Z$ parameter vs temperature for devices A and B](image)

Figure 5.23: The “$Z$” parameter at threshold plotted as a function of temperature over the range 40 K - 80 K for devices A (black) and B (blue).

<table>
<thead>
<tr>
<th>Structure ref</th>
<th>$J_{\text{def}}$ (%)</th>
<th>$J_{\text{rad}}$ (%)</th>
<th>$J_{\text{leak}}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (40 K)</td>
<td>82 ± 4</td>
<td>18 ± 2</td>
<td>0 ± 6</td>
</tr>
<tr>
<td>A (80 K)</td>
<td>78 ± 4</td>
<td>9 ± 2</td>
<td>13 ± 6</td>
</tr>
<tr>
<td>B (40 K)</td>
<td>68 ± 4</td>
<td>32 ± 2</td>
<td>0 ± 6</td>
</tr>
<tr>
<td>B (80 K)</td>
<td>66 ± 4</td>
<td>27 ± 2</td>
<td>7 ± 6</td>
</tr>
</tbody>
</table>

Table 5.2: The approximate threshold component fractions at 40 and 80 K for structures A and B.

The percentages of recombination associated with each mechanism are tabulated in table 5 for both structures at 40 K and 80 K. Firstly, as was seen in section 5.2 SRH recombination is clearly the dominant recombination mechanism in both devices and therefore the main efficiency limiting mechanism. The increased SRH recombination of device A compared to device B implied from the characteristic temperatures and $Z$ analysis is confirmed explicitly and quantified here to be 14% at 40 K and 12% at 80 K. This increase in defect density is attributed to a less optimal Si surface orientation causing reduced double atomic step generation at the Si
surface. As a result additional strain and APDs are introduced in the epitaxially deposited III-V material, leading to threading dislocations that may reach the active region as was discussed above.

The defect density, although lower in the case of device B is still unacceptably large, accounting for the majority of injected carrier recombination. Therefore it is possible that the Si surface orientation chosen for device B is still not quite at the optimal value and could potentially still be improved. However, the analysis carried out in section 5.2 indicates that nitrogen diffusion may also be responsible for the generation of additional defects at the GaNAsP/GaP interface. Therefore Si surface orientation is likely only one of several sources of defects, limiting the degree to which its optimisation can reduce SRH recombination alone.

The leakage component at 80 K is also found to account for 6% more of the total threshold current in device A. However, based on the error associated with the approximation of the carrier leakage fraction using this method this difference cannot be considered statistically significant from this data alone.

![Chart](chart.png)

Figure 5.24: The defect, radiative and carrier leakage components of the threshold current density plotted as a function of temperature for devices A (black) and B (blue). Based on multiple sources of evidence for the presence of carrier leakage the leakage component is determined as: $J_{\text{leak}} = J_{\text{th}} - J_{\text{def}} - J_{\text{rad}}$.

### 5.3.3 Pressure dependent electroluminescence

The application of high hydrostatic pressure to reversibly alter the band gap offers an alternative, complimentary route to investigate the recombination mechanisms of these devices. The methodology and experimental arrangement here is identical to that of section 5.2, allowing direct comparison of results. By measuring the lasing wavelength as a function of pressure the pressure coefficient of the band gap can be determined for each device, as is depicted in figure 5.25 for a temperature of 80 K. A slight superlinearity to the data is thought to be caused by an increase in $J_{\text{th}}$ with pressure, leading to a small blue shift in emission wavelength.
(increase in energy) since a higher injection current is required to reach $1.1 \times J_{th}$.

Devices A and B are found to have pressure coefficients of $61 \pm 2 \text{ meV GPa}^{-1}$ and $62 \pm 2 \text{ meV GPa}^{-1}$ respectively, matching that of #16052 with a value of $61 \pm 2 \text{ meV GPa}^{-1}$. All values are within the margin of error for the experiment and since the only difference expected between these devices is the surface orientation this matches expectations of a near identical active region composition (other than defect density). With this confirmed we can state that the activation energy of the carrier leakage path into the X-minima would match that of #16052 with a value of $\sim 71 \text{ meV GPa}^{-1}$ and that the subsequent analysis which found this leakage path to be minimal would hold in this case also.

![Graph](image)

Figure 5.25: Lasing energy plotted as a function of hydrostatic pressure for devices A and B at 80 K.

Figure 5.26 (left) shows the pressure dependence of the normalised threshold current density at 80 K for devices A and B. From this plot we see that device B has a much stronger pressure dependence than device A, although still far from the ideal value expected for a device dominated by radiative recombination (red line). The ideal radiative plot is determined from figure 5.25 in conjunction with the expected dependence of the radiative threshold current density on band gap, $J_{rad} \propto E_g^2$. The $J_{th}$ of a device dominated by defects is expected to be almost completely independent of pressure since defect density is the dominant factor in the SRH recombination rate and this clearly remains unchanged, therefore the horizontal brown line represents this limiting case. The normalised plot causes the trend in $J_{th}$ with pressure to look considerably stronger than the radiative limit, suggesting a relatively high carrier leakage components, however plotting on an absolute scale (figure 5.26, right) shows that this is in fact not the case.

The theoretical behaviour of a leakage dominated device was determined assuming the leakage path pressure coefficient of approximately $58 \text{ meV GPa}^{-1}$ for #16052 and is plotted in green. Since we have seen
strong evidence from the temperature dependence study as well as the previous analysis in section 5.2 that these devices are limited mainly by defect related recombination and carrier leakage there are again two possible explanations for the increased pressure sensitivity exhibited by device B in figure 5.26.

Either the leakage fraction of device B has been increased, which would lead to reduced performance or the defect fraction has been decreased, leading to an improvement in performance. Clearly the latter must be the case given the superior threshold current density and characteristic temperatures as well as the quantified threshold components from figure 5.24. The decrease in SRH recombination for device B causes the pressure dependence of \( J_{th} \) to increase well beyond the ideal radiative value, indicating that although defects are the main limiting factor carrier leakage is far from negligible and would also have to be addressed to achieve an efficient laser operating at non cryogenic temperatures.

Figure 5.26: Left: Normalised threshold current density as a function of hydrostatic pressure as well as theoretical 100% defect, radiative and carrier leakage (brown, red and green respectively) limits for reference at 80 K. Right: As left but on an absolute scale.

### 5.3.4 Conclusions

The performance and efficiency limiting processes of two GaNAsP/Si single quantum well laser structures based on structure #16052 were investigated as a function of Si surface orientation. Devices A and B were both grown with a 0.1° miscut from the (001) but device A additional has a 30° miscut away from the (110) direction (rotation in the [001] plane). The results of the temperature and pressure dependent electroluminescence investigation are summarised in table 6 below.

<table>
<thead>
<tr>
<th></th>
<th>( J_{th}(kAcm^{-2}) )</th>
<th>Max operating temp (K)</th>
<th>( T_0(K) )</th>
<th>( T_1(K) )</th>
<th>( Z_{th} )</th>
<th>( J_{def}(%) )</th>
<th>( J_{rad}(%) )</th>
<th>( J_{leak}(%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device A</td>
<td>3.29</td>
<td>80</td>
<td>45 ± 3</td>
<td>33 ± 10</td>
<td>1.39 ± 0.04</td>
<td>78 ± 4</td>
<td>9 ± 2</td>
<td>13 ± 6</td>
</tr>
<tr>
<td>Device B</td>
<td>0.87</td>
<td>120</td>
<td>119 ± 3</td>
<td>60 ± 10</td>
<td>1.65 ± 0.04</td>
<td>66 ± 4</td>
<td>27 ± 2</td>
<td>7 ± 6</td>
</tr>
</tbody>
</table>

Table 5.3: Summary table for surface orientation dependent parameters at 80K.
The reduction in $J_{th}$ from $3.29 \text{ kAcm}^{-2}$ to $0.87 \text{ kAcm}^{-2}$ and increase of operating temperature from 80 K to 120 K, comparing device A to B alone clearly indicate that device B is the superior of the two in performance but gives no indication as to why. The low $T_0$ value of device A is reasonably close to the theoretical limit for a defect dominated device of $(\frac{2T}{3})$, far less than the value for device B, which is indicative of an increased defect density within device A.

The $T_0$ value of device B exceeding $T$ is attributed to the inhomogeneity of the active region as discussed in section 5.2. The fairly low $T_1$ values of 33K and 60 K for devices A and B respectively are suggestive of high internal losses (since little evidence of optical losses has been found) in both devices, but to a greater extent in device A. A reduced internal loss for device B is consistent with a reduction in defect density and therefore scattering cross section within the device. The Z analysis also indicates reduced defect related recombination in device B compared to A since values of 1 and 2 represent pure SRH and radiative recombination respectively.

Separating the threshold current density as a function of temperature into components using the method detailed in section 5.2.1 it was found that the defect current contribution was 12 % lower in the case of device B, the radiative current fraction was 18 % higher and the leakage fraction was 6% lower. These values back up the other findings suggesting that the defect density is greater in device A and explains the associated reduction in maximum operating temperature and increase in threshold current density. In addition pressure dependent electroluminescence supported these findings by indicating that the normalised threshold current density pressure dependence of device A is close to the limiting case of a completely defect dominated device.

The overall conclusion of this study is therefore that firstly device B clearly out performs device A in all aspects and therefore has the superior substrate orientation. Secondly the difference in performance appears to be tied almost exclusively to a reduction in defect related recombination and therefore a reduction in defect density within the active region. Based on the background theory and previous studies discussed at the beginning of this section it is reasonable to conclude that this reduction in defect density is the result of achieving a Si surface orientation that is closer to optimal in the case of device B and that this has led to greater double step formation, as is desired for defect free growth.

Although device B out performs device A defect related recombination remains one of the dominant efficient limiting processes in the GaNAsP/Si based devices that requires further reduction in order to achieve a commercially viable product. This could be as a result of the Si surface orientation of device B being better optimised but not yet having reached an ideal value. If this is the case, then as this investigation has shown, even a small change in orientation can yield a substantial increase in performance and so further study would be advised in order to achieve the ideal orientation. In addition the study presented in section 5.2 has indicated that there are at least two plausible sources of defects in these devices (those induced by strain and APDs as well as potentially those caused by the diffusion of nitrogen). If this is the case then lowering SRH recombination will require addressing several independent issues and eliminating each source individually.
6 Simulation of the GaNAsP/BGaP/Si material system

Following the study of the GaNAsP/BGaP/Si based lasers of structures #16052, #16053 and #16195 it was found that #16195 exhibited the greatest performance. Accordingly, subsequent growth runs produced both a reproduction of the #16195 structure, under the same conditions as the original run as well as alternative structures in order to explore further routes of optimisation. However, all structures received after those investigated in chapter 5 were found not to lase at any temperature and only pWs of emission power were collected from the best performing samples. The number of devices even producing spontaneous emission was found to be minimal from batches of ~50-100 devices, with the majority “burning out” and ceasing to conduct a current beyond comparatively low current densities of $1 \text{ kAcm}^{-2}$.

Since each batch always contained a copy of the #16195 structure the failure of all devices regardless of structural or growth properties was an indicator of an external issue affecting all devices. The MOCVD reactor used for the growth of these devices was in continuous use as a part of a range of projects in between the growth of GaNAsP based materials with shared precursors, which were changed several times over the period devices were received. Other devices grown in different material systems as a part of other projects were not found to exhibit the same drop in performance and so a fault with the reactor and/or contamination of the precursors was ruled out.

The only factor known to have changed between the growth of functional and non-function devices was the source of the Si substrates on which the GaNAsP/BGaP/Si lasers were grown. The original Si substrates were exhausted and it was found that the product had been discontinued, therefore an alternative supplier was found from which to source substrates of an identical specification. Ongoing investigation at Marburg, including TEM and AFM imaging has indicated an increased formation of APBs in the GaP nucleation layer grown according to the conditions previously found to be optimal.

It is thought that a difference in surface quality or morphology is related to the reduced performance of devices grown on this substrate. Since the studies carried out at the University of Surrey are focused on device characterisation and optimisation rather than growth it is challenging to produce meaningful data from non-functional lasers. As a result the material system was investigated from a theoretical perspective by computer simulation of optical properties, as well as a study of device processing methodology. In doing so it was anticipated that progress could be made in terms of optimisation of the device structure and processing techniques whilst the growth issues were resolved separately.

6.1 Optical simulation of single and multi-quantum well GaNAsP/Si laser structures

Through study of the documentation of GaNAsP based laser development it was determined that the thicknesses of layers and material choices had mainly been considered in terms of minimising lattice mismatch,
strain and defect generation. Having seen how vital the minimisation of defect density is in these devices this approach is logical. However, equations 2.32 and 2.21 concerning the relationship between $n_{th}$, gain and $I_{th}$ highlight the importance of maximising the optical confinement factor. Since $I_{th}$ and $n_{th}$ are exponentially dependent on the modal gain, a moderate increase in optical confinement factor will lead to disproportionate improvements in overall device performance. If the growth properties require re-optimisation in order to once again achieve lasing, additional improvements through maximisation of the optical confinement factor will be beneficial in achieving this goal.

6.1.1 Waveguiding theory

Maximising the confinement factor is a case of making the core, i.e. the quantum wells and barriers as good a waveguide as possible relative to the cladding material. How strongly the light couples into the waveguide is related to two key factors, the refractive index difference between the core and the cladding ($\Delta n$), as well as the thickness of the active region. Wave guiding takes place through total internal reflection (TIR) of photons produced in the active region at material layer boundaries where there is a refractive index step. The angle of transmission, $\theta_2$ of an incident wave at an angle, $\theta_1$ to a boundary between two media of refractive indices $n_1$ and $n_2$ is expressed through Snell’s law as follows:

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2).$$

(6.1)

For wave guiding we require total internal reflection, meaning in practice we require $\theta_2 \geq 90^\circ$, which from equation 6.1 requires a critical angle of incidence ($\theta_c$) that can be expressed in the form of equation 6.2,

$$\theta_c = \arcsin\left(\frac{n_2}{n_1}\right).$$

(6.2)

For equation 6.2 to hold it is required that $n_1 > n_2$, i.e. that the material in which the wave is to be confined has a refractive index exceeding that of the surrounding material. Meeting this requirement in the case of a laser is straight-forward since the refractive index is inversely proportional to the band gap, and the quantum well and barrier band gaps will always be less than that of the cladding in order for carrier injection to be possible. Evaluating equation 6.2 for the barrier/cladding and the QW/barrier interfaces of the GaNAsP based lasers gives critical angles of 72° and 81° degrees respectively. The determination of the relevant refractive indices is discussed in section 6.1.2. Since lasers by nature are highly directional the critical angle requirement should be easily met for these photons, suggesting that confinement should be near perfect.
Figure 6.1: The laser geometry assumed in the mathematical treatment of the optical mode below. X, Y and Z are taken to be the cavity, growth and transverse directions respectively.

However, the confinement of photons within a waveguide is somewhat analogous to the confinement of electrons within a potential well and the mathematics is therefore somewhat similar. As such, just is the case with the finite well potential a confined mode will extend to some extent into the surrounding material even if TIR takes place. The quantification of this overlap, known as the evanescent field/mode, will be required to calculate the spreading of the optical mode across the structure in a simulation. The photons propagating along the axis of the cavity (x) are described by a plane wave equation (6.3),

\[ E = E_y(x) \exp(i(\beta z - \omega t)) \quad (6.3) \]

where it is assumed that the mode has no dependence on the transverse direction (y). Solving for the confined modes again follows an analogous treatment to the finite well problem, the details of this can be found in appendix B along with the derivation of the so called V-parameter. The V-parameter, given by equation 6.4,

\[ V = \left( \frac{k_0 h}{2} \right) \left( n_1^2 - n_2^2 \right)^{1/2} \quad (6.4) \]

can be considered to be related to the degree of confinement of the mode within the core region. Since the modes decay exponentially into the cladding they theoretically have infinite extent and so a consistent measure for the extent of the optical field is required. The mode field diameter (MFD) is defined as twice the distance from the centre of the core (for a symmetric wave guide) to the point at which the field strength drops to 1/e. The functional form of the MFD is defined as:

\[ MFD = h + \frac{2}{\gamma}. \quad (6.5) \]

\(\gamma\) is the wave vector of the lowest order confined mode. The MFD may also be approximated as:

\[ MFD \approx h \left( \frac{V + 1}{V} \right). \quad (6.6) \]
There are two regimes to the MFD, for a core thickness of several \( \lambda \) or more it scales with \( h \), essentially the core thickness plus the extent to which the evanescent tail of the mode extends into the cladding on either side. The penetration of the evanescent part of the mode will be exponentially dependent on the absorption coefficient of the cladding.

However, for waveguides of core thickness less than the wavelength of the photons to be guided then the MFD will increase rapidly as a result of two additional factors. The first is simply that the field within the core will be diffraction limited, i.e. \( MFD > \frac{\lambda}{2\pi} \). Secondly as the waveguide becomes smaller the strength of the guiding is reduced since an increasingly large proportion of the mode will be accounted for by the evanescent overlap into the cladding. This can be seen from equation 6.5; as \( h \) tends to zero the fraction of the mode that is outside the core region tends to 100%.

This is the reason one may not simply confine the optical mode within a quantum well, a few nm in thickness and instead a SCH structure is required to achieve reasonable confinement. For the case of the GaNAsP material system the QW has a thickness of 5 nm, but the diffraction limit means the smallest the MFD could ever reach would be \( \sim 110 \) nm. Using equation 6.6 this value increases to \( \sim 340 \) nm, slightly larger than the thickness of the barriers and QW combined for #16195.

Therefore, we see that the degree of confinement is not only dependent on the magnitude of the index discontinuity but also the width of the waveguide core. Considering the QW and barriers as the core the MFD value rises to \( \sim 890 \) nm, considerably larger than the core diameter, which is nominally 311 nm for #16195. The confinement factor, discussed in section 2.6.3 of chapter 2 relates directly to the MFD, since the minimisation of the optical field spread will maximise its amplitude at the quantum well.

### 6.1.2 Simulation details and methodology

The structures used as a starting point for the optical simulation were the best performing single and multiple quantum well structures, namely #16195 (figure 5.7) and #16053 (figure 5.8). Section 6.1.1 demonstrated that the only parameters required for a one dimensional waveguiding simulation are the refractive indices of the materials and the layer geometry. Generally the refractive indices of most commonly used semiconductors and their alloys are well known and documented [102].

In the case of GaNAsP, BGaAsP and BGaP however, these are not readily available due to the relative lack of detailed study, particularly as a function of composition for these materials. Thus for the purposes of this simulation some approximations are made in order to define suitable values for the refractive indices of these novel materials. Work carried out at the Fraunhofer institute, Freiburg, in collaboration with the University of Marburg presents ellipsometry data for both BGaP and BGaAsP at various compositions as well as GaP [136]. Extrapolation to the composition and lasing energy relevant to this study allows an approximation of the refractive indices of these materials to be made. The presence of GaP data within the ellipsometry study allows a comparison to be made between the extrapolation and the well known and documented refractive index of GaP. It is found that the extrapolated GaP value differs from that presented in the literature [145] by only 0.005, suggesting that the extrapolation may be reasonable. For the active
material, GaNAsP similar studies have not been carried out to determine the refractive index. Thus equation 6.7[143], which relates this to the bandgap is used as an alternative approximation:

\[ n = K E_g^C. \] (6.7)

Here \( K \) and \( C \) are experimentally determined constants and are found to be equal to \( 3.668 \, eV^{-1} \) and -0.32234 respectively, as reported in [143]. The bandgap is determined from the experimentally measured lasing energy, taking into account that this emission will result from the \( E_1 \) state of the quantum well rather than the bulk bandgap. The approximate difference can be determined using the estimate of the nitrogen concentration in the quantum well produced in chapter 5 (6 - 6.9 %) along with the band anticrossing model.

Although this approximation is fairly rough, in section 6.1.1 it was shown that the effect of the quantum well on confinement is minimal since \( \lambda \gg \) thickness of this layer. The values determined for the refractive indices of BGaP, BGaAsP and GaNAsP used in this simulation, based on the aforementioned approximations are summarised in table 7.

The simulations were carried out in the LaserMod application from the Rsoft (now Synopsys) software package, details of which can be found in [149]. LaserMod is capable of carrying out a wide range of simulations, from passive waveguiding all the way up to electrical injection lasing. In order to carry out a simulation the structure under investigation must first be defined using a CAD style interface. For the purpose of the waveguide simulation only a two dimension cross-section is required, although a full three dimensional model may be defined in order to take into account cavity effects at a later stage as required. This however comes at the cost of increased demand on system resources.

The layers are then assigned materials and electrical contacts are placed on the N and P sides. All layers have their thicknesses and positions in the structure defined as variables, allowing them to be varied by a script where necessary. Layers are also defined as either “bulk semiconductor” or “active”, allowing the simulation to differentiate between quantum wells and barrier regions for example. A grid is then applied over the structure (a finite element method is used), which determines how many and at which positions calculations are carried out. This is depicted in figure 6.2, the red lines indicate layer interfaces and the black lines the grid.

For example, if the grid spacing were set to 1 nm, Maxwell's equations would be evaluated once per nm in a regular square pattern across the device. Thus, a finer grid gives a closer approximation to the solution of the equations but increases run time and so these two factors require balancing. The spacing was set to vary, shorter near interfaces and in thinner layers and larger in thick layers, far from interfaces. This allows an accurate representation of the optical field to be simulated without consuming excessive system resources.
When defining a material for each layer all material parameters must be included in order for the simulation to run, regardless of whether they are required. This was an issue as the novel Boron and Nitrogen containing materials used in these devices do not have well documented properties that could be used to fully define the materials within LaserMod. The refractive indices mentioned above were matched to materials within LaserMod’s existing database that could be substituted into the structure. The substituted materials and the approximation of their refractive indices are shown below in table 7.

The fact that the materials are substituted should make no difference to the simulation, provided only the optical mode and confinement factor are calculated, since the only material parameter called is the real part of the refractive index. These substitutions do make further in depth simulations using this material system problematic however. Therefore for full device simulation the remaining unknown material parameters would need to be either experimentally determined or at least approximated after a relevant theoretical study.

<table>
<thead>
<tr>
<th>Original material</th>
<th>BGaP</th>
<th>BGaAsP</th>
<th>GaNASP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substitution</td>
<td>GaAs$<em>{0.47}$P$</em>{0.53}$</td>
<td>GaAs$<em>{0.62}$P$</em>{0.38}$</td>
<td>$In_{0.083}Ga_{0.917}As$</td>
</tr>
<tr>
<td>Refractive Index</td>
<td>3.154 ± 0.005</td>
<td>3.195 ± 0.005</td>
<td>3.318 ± 0.08</td>
</tr>
</tbody>
</table>

Table 6.1: Substituted materials and their refractive indices.

The mode calculation carried out by LaserMod uses a Ritz iteration method to solve Maxwell’s equations in one dimension (the growth direction) across the layer interfaces. The Ritz method is a form of finite element method, and therefore requires the problem to be broken down into parts of a finite size, this is the job of the aforementioned grid. Details of the Ritz method can be found in reference [127], with an example of its
use in laser development shown in [23]. The confinement factor is then determined by summing the optical field intensity over all layers set as “active” (the QWs) and dividing by the sum of the optical field intensity over the full extent of the structure. A more detailed explanation is presented in appendix C.

6.1.3 Single quantum well structure optimisation

Initially structure #16195, having yielded the best performing single quantum well devices was produced using the CAD interface, and the refractive index profile and optical mode simulated. Figure 6.3 depicts the optical mode and refractive index profile in the growth direction across the whole III-V region at the top and zoomed to the active region on the bottom. Several unexpected features of the optical field can be seen from figure 6.3. Firstly the quantum well position (marked by the red dotted line) is offset from the peak of the optical mode. The mis-alignment is found to be ∼110 nm, more than 22 times the thickness of the quantum well, causing a drop in field strength of ∼8%.

The confinement factor is found to be only 0.35% as a result of the dispersed and mis-aligned optical mode, a strong indication that there is significant scope for improvement with this structure. The structure in the immediate vicinity of the quantum well is fairly symmetric and should not cause such an offset of the optical mode. However, the field strength at the BGaAsP capping layer is also significant, even through it is separated from the P-side barrier by 1.2 µm of BGaP cladding.

For the mode to be coupled into the capping layer the confinement in the active region must be particularly weak, which is consistent with the low refractive index contrast between the BGaAsP barriers and BGaP cladding. The result is that whilst total internal reflection (TIR) takes place on the n-side of the structure, on the p-side frustrated internal reflection (FIR) is induced. This phenomena takes place when the evanescent part of the mode reaches a material with a greater refractive index than the cladding. Usually the spreading of the evanescent mode is only problematic in that the Gaussian mode is spread over a larger area, reducing the peak field strength, since no energy is transmitted by the evanescent mode.

However, in the case of frustrated internal reflection energy may be transmitted across the cladding and into the higher index material on the other side. This process is analogous to the quantum tunneling of charge carriers from a potential well in quantum mechanics. With reasonable coupling into the capping layer the structure acts as if there are two waveguides of reasonably equivalent strength leading to the photons being shared between both and the field becoming offset from the quantum wells. It should be noted that optical absorption shouldn’t be too much of a problem outside of the active region since the lasing energy is significantly less than the bandgaps of the cladding and capping layers.
Figure 6.3: Simulated optical field and refractive index for structure #16195 as a function of structure thickness measured in the growth direction. Zero marks the top surface of the Si substrate and the red and blue dotted lines mark the QW and significant material boundaries respectively. $\Gamma = 0.35\%$. QW/optical field mis-alignment = 110 nm. Top: Full device. Bottom: Zoomed to active region.

With reference to figure 6.3 two issues have been identified, which if addressed should improve the optical confinement factor. Firstly, by simply increasing the distance between the active region and the BGaAsP capping layer the leakage of the transverse mode into this region can be almost entirely eliminated. This can be achieved by increasing the thickness of the barrier and/or cladding regions on the P side of the device. In doing so this should also reduce the asymmetry of the mode and reduce the QW/peak position mis-alignment.

Secondly the aforementioned mis-alignment should be reduced to as close to zero as possible in order to maximize the photon density within the quantum well. This can be primarily achieved by increasing the thickness of the barrier regions on both sides, large enough that the waveguide strength is improved but not so far that the mode is spread over a large region, reducing the peak intensity. The n-side cladding region thickness is expected to have less influence on the confinement factor as there are no higher index steps on the n-side of the device and as such the optical field can be seen to tend to zero.

The first optimisation investigation consisted of varying the n and p-side cladding (BGaP) layers individually from 1 to 5 $\mu$m and 0 to 5 $\mu$m respectively in steps of 25 nm. The n-side BGaP cladding is not simulated below 1 $\mu$m since this is approximately the minimum value for which the laser will function as a result of several factors aside from optical confinement. Carrier injection via the lateral contacts takes place through the n-BGaP, requiring a reasonable thickness to act as a lateral current channel. In addition this layer helps to isolate the active region from the III-V/Si interface, at which APDs, threading dislocations and
other growth defects may be present. The confinement factor was evaluated for each increment of p-BGaP and n-BGaP cladding thickness and plotted below in figure 6.4 as the black and blue data respectively.

As is to be expected, varying the n-BGaP layer thickness has negligible impact on confinement factor since there is no higher index material beyond the n-BGaP layer for the optical mode to interact with. The p-side BGaP layer thickness however has a strong impact on confinement factor, with a maximum found at either extreme of the layer thickness.

With reference to the discussion of coupling the optical mode in the capping layer above it is logical that an increase in confinement factor is observed with increasing p-BGaP thickness, asymptoting towards \( \sim 0.48\% \). The coupling of the mode into the capping will decrease as the separation is enlarged, increasing both the field strength at the “primary peak” of the mode around the active region as well as re-centering the mode peak on the QW. The asymptotic nature of the curve leads to diminishing returns in confinement factor for increased thickness and so a cut off value of \( \sim 4.5 \mu \text{m} \) was chosen beyond which further thickness increase was of no value.

As the p-BGaP thickness is reduced from the original value the confinement factor rapidly increases, having more than doubled below 200 nm. On first inspection the implication here is that removing the p-BGaP layer entirely would lead to the best confinement factor and therefore lasing performance. When considering why the confinement factor would increase with decreasing p-BGaP thickness it can be seen from 6.3 that this would bring the BGaAsP capping layer closer to the active region, increasing the coupling into this region.

Beyond the BGaAsP capping layer is air, which with a refractive index of 1 causes very strong guiding along this interface. This means that the mode will not extend far beyond the capping layer but since the region is fairly thin (100 nm) the guiding is still fairly poor. Thus the offset of the primary mode from the active region is being reduced as the two peaks are pushed together and cannot leak further beyond the capping layer. As the p-BGaP layer thickness tends to zero the barriers and capping layer are very strongly coupled and the effective cladding material which confines the mode becomes the air beyond the capping layer. As a result the confinement on this side of the active region is very strong with a large refractive index contrast in excess of 2, compared to 0.04 previously. Thus the fundamental cause of the rapid increase in confinement factor with decreasing p-BGaP thickness is the guiding of the optical mode against the capping/air interface.

For the confinement factor to exceed the \( \sim 0.48\% \) maximum value achieved by increasing the p-BGaP thickness a value below \( \sim 400 \text{ nm} \) would be required. There are several reasons why such a value would be unsuitable for producing a reliable laser. Firstly, the simple fact that the active region would be within \( \sim 500 \text{ nm} \) of the surface of the device means there is a reasonable chance of damage from even fairly shallow scratches during wafer and device processing as well as integration steps. In addition the annealing of the p-contacts causes the diffusion of Zn dopants in order to form a continuous doping profile and thus produce good Ohmic contacts. Thus, a close proximity of the active region to the contacts can cause Zn defects to reach the barriers, or worse the QW where they can create states into which IVBA can take place. Also, when
integrated onto Si the lasers will likely be buried within the surrounding material and thus the BGaAsP/air interface which is the cause of this peak in confinement factor will cease to exist. Worse still, if the surrounding material is Si then the refractive index contrast will actually become negative, causing anti-guiding. Thus it is unwise to design a device which relies on the surroundings to operate as intended. The result is choosing to increase the p-BGaP thickness in order to achieve optimisation rather than decrease it.

Figure 6.4: Confinement factor Vs \( p \) and \( n \) (blue and black) BGaP cladding thicknesses. In each case the other layer thicknesses are set to their original values. The values for the n and p BGaP used in \#16195 are marked by the dotted vertical lines.

Plotted below in figure 6.5 is the optical mode and refractive index profile for structure \#16195 with the p-BGaP cladding thickness increased to 4.5 \( \mu m \). As was discussed above with reference to figure 6.4 it can be seen that the increase in p-BGaP thickness has moved the capping layer far enough from the active region to prevent FIR producing the double peak observed in figure 6.3. All photons are now contained in the primary peak, increasing the peak field strength and causing a shift towards the centre of the active region, reducing the misalignment to only 13 nm from 100 nm.

The result is an increase in optical confinement factor from 0.35% to 0.48%, which as a fractional increase is \( \sim 27\% \). The large increase in device thickness, although beneficial in terms of optical confinement would also be detrimental in some aspects. The increased material consumption is one factor, simply in terms of making the devices economically and eventually achieving a suitably priced component. More practically though are the implications for carrier injection and electrical performance. A thicker device increases the distance carriers must travel before reaching the QWs, giving more opportunities for scattering and monomolecular...
recombination, increasing the series resistance of the device and potentially causing further heat dissipation issues at high injection. In practice therefore, we would not suggest such a large value for the p-BGaP thickness, however as a first step in the optimisation it is helpful to fully isolate the optical mode from the capping layer.

After the BGaP cladding the other layers to be optimised are the BGaAsP barriers. From equations 2.30, 6.4 and 6.5 it can be seen that there is a balance to be made between core thickness and confinement factor at the QWs. Raising the core thickness will increase both V and the MFD, however at some point the mode will be spread over such a large core that the peak intensity will start to drop, lowering the optical confinement factor. For the second batch of simulations both BGaAsP barriers were varied independently of each other over the range 0 to 1 µm in steps of 50 nm.

A script written in MatLab first set the n-side barrier thickness to 0 µm by editing the LaserMod structure text file, it would then call the parameter scan function of laser mod to vary the p-side barrier thickness over the aforementioned range. The confinement factor is read off from the output text file along with the optical field and index profiles and the n-side barrier thickness incremented by 50 nm and the process repeated. Figure 6.6 was plotted such that the x-axis denotes the thickness of the p-side barrier and the y-axis the confinement factor.

Each data set represents a different n-side barrier thickness, as denoted in the legend. The black data set

Figure 6.5: Simulated optical field and refractive index for structure #16195 with a p-BGaP cladding thickness of 4.5 µm. Γ=0.48%. QW/optical field mis-alignment = 13 nm. Top: Full device. Bottom: Zoomed to active region.
represents the standard n-side barrier thickness for structure #16195 (150 nm) and the blue data the value which gave the largest peak confinement factor (300 nm). In each case the circled data point indicates the associated p-side barrier thickness for the standard (black) and optimised (blue) structures.

For decreasing values of barrier thickness the confinement factor drops off rapidly, as expected due to the weakening of the waveguide, since the majority of the mode becomes accounted for by the decaying tail in the BGaP cladding. An increase in barrier thickness leads to a peak value for both sides of 300 nm, after which the confinement factor begins to drop off as a result of the mode being spread over an increasingly large area. Increasing the p-side barrier thickness alone to 300 nm yields an increase in confinement factor from 0.48% to 0.56% with a further improvement to 0.6% attained when the n-side thickness is increased also. The reason for the difference in improvement between the n and p barrier side barrier changes is likely the result of the asymmetrical mode, which is slightly offset from the QW.

Figure 6.6: Confinement factor Vs p-barrier thickness for the standard #16195 n-barrier thickness of 0.15 \( \mu m \) (black) and optimised value of 0.3 (blue). The marked data are the standard #16195 confinement factor (black) and optimised for both barrier thicknesses (blue).

Figure 6.7 show the optical mode and refractive index profiles for the structure with optimised barrier thicknesses (0.3 \( \mu m \)). There are three clear improvements in comparison to the previous profiles. Firstly the mode peak misalignment from the QW has decreased to 3 nm, which gives an almost undetectable drop in optical field strength at the quantum well. Secondly, the optical mode is much narrower than previously. Visual comparison shows a drop in separation of the point at which the mode appears to drop to zero from 4 \( \mu m \) to 3 \( \mu m \) as a result of the barrier optimisation. Finally, the mode amplitude increases from \(~5.2\) to \(~5.8\), the scale is in arbitrary units but the fractional increase of \(~10\%\) is still valid and will be the main
factor attributed to the increase in optical confinement factor.

Figure 6.7: Simulated optical field and refractive index for structure #16195 with a P-BGaP cladding thickness of 4.5 \( \mu m \) and barrier thicknesses of 0.3 \( \mu m \) zoomed to active region. \( \Gamma = 0.6\% \). QW/optical field misalignment = 3 nm. Top: Full device. Bottom: Zoomed to active region.

Having optimised the cladding and barriers once each the cladding optimisation was repeated to determine if the original optimised value could be somewhat reduced due to the improved confinement resulting from the barrier optimisation. Reducing the p-BGaP thickness from the particularly large value of 4.5 \( \mu m \) down to 1.8 \( \mu m \) had no effect on confinement factor or coupling into the capping layer with the optimised barriers. This brings the p-BGaP thickness much closer to the initial value of 1.2 \( \mu m \), minimising the impact on carrier injection and electrical properties of increased device thickness but maintaining maximal optical confinement factor.

The resulting final structure is shown in figure 6.8 (bottom) below along with the standard structure for comparison (top). When compared to the original field profile there is a considerable improvement, with an optical confinement increase from 0.35\% to 0.6\% and a QW/optical mode misalignment decrease from 100 nm to 3 nm. A small improvement in confinement factor could be achieved by using smaller increments in barrier and cladding thickness, however the resulting change in confinement factor would be less than 0.005\%. Due to the previously mentioned approximations required to carry out the simulations, optimising to such precision would be somewhat meaningless as a result of the uncertainty in some of the material parameter values.

The reduction in threshold gain achieved by increasing the optical confinement factor from 0.35\% to 0.6\%
can be approximated from equation 2.28 to be 42% if one assumes constant internal and mirror losses. From equation 2.29 this assumption holds for the mirror losses, however the internal optical losses will clearly vary to some extent as the field strength and percentage of it that propagates in each layer has changed. Therefore this approximation determines the best case scenario but gives a rough indication of the reduction in gain that could be expected.

Expanding the approximation to the threshold carrier density using equation 2.32 gives a reduction of 35% and then substituting into 2.21 along with the recombination mechanism fractions from chapter 5 for structure #16195 (52%, 31% and 17% for SRH, radiative and carrier leakage respectively) give a drop in $I_{th}$ of 41%. The drop in $I_{th}$ in addition to the previous assumptions also requires that the increase in optical confinement factor does not alter the recombination rates of each mechanism, which again is not entirely physical. The radiative rate will increase since stimulated emission is dependent on the photon density, whilst the SRH rate should remain constant and carrier leakage decrease.

Taking into account that the above approximations are fairly loose the values themselves should not be treated as representative of the reality of the situation. However, the 41% drop in $I_{th}$ if even a quarter of this value in actuality would still represent a very significant improvement in device performance.

A problem with this material system is that the refractive indices of GaAs and GaP are $\sim 3.3$ and $\sim 3$, meaning that even if the barriers and cladding were pure GaAs and GaP the index step would still not be that large. Therefore small changes in composition, that may not too detrimentally effect other parameters such as lattice constant and band gap too significantly will have a fairly minimal impact on refractive index contrast. For example, the barrier composition used in #16195 is $B_{0.05}Ga_{0.11}As_{0.89}P$, even doubling the As content to 20% would only increase the optical confinement factor from 0.6% to 0.7%. These values are calculating by evaluating equations 6.4 and 6.6 with the relevant refractive index at the new composition and then determining the optical confinement factor for that MFD.
Figure 6.8: Simulated optical field and refractive index for structure #16195. Top: Standard structure, $\Gamma = 0.35\%$. QW/optical field mis-alignment = 110 nm. Bottom: Optimised with a P-BGaP cladding thickness of 1.8 $\mu$m and barrier thicknesses of 0.3 $\mu$m. $\Gamma = 0.6\%$. QW/optical field mis-alignment = 3 nm.

6.1.4 Multiple quantum well structure optimisation

Having carried out an initial optimization of device #16195 it seemed logical to also do the same for device #16053 as the idea of this structure was to increase confinement factor through the inclusion of additional quantum wells. One would expect similar limiting factors for this structure since it is very similar to #16195, all materials and compositions are identical, only layer thicknesses and the structure of the active region differ.

Figure 6.9 depicts the optical field and refractive index profiles for structure #16053; in this case the optical confinement factor is 1.33\% and the QW/optical field misalignment is 94 nm. Generally one would expect that a tripling of the number (and therefore total thickness) of QWs would yield approximately a factor of three increase in optical confinement factor, however in this case the factor is closer to 4. This increase in confinement factor has already been shown to lead to an improved device performance through the work presented in [111]. The larger than expected increase in optical confinement factor is through to result from two factors. Firstly the total thickness of the active region increases from 308 nm in the case of the single QW structure, #16195, to 404 nm for this structure due to the additional barrier material required to separate the additional QWs. Secondly, the mis-alignment difference between #16195 and #16053 as well as the resulting dissimilar optical confinement factors of the three wells causes the sum to be larger than expected.
Figure 6.9: Simulated optical field and refractive index for structure #16053 (MQW) as a function of structure thickness measured in the growth direction. $\Gamma = 1.33\%$. QW/optical field mis-alignment = 94 nm. Top: Full device. Bottom: Zoomed to active region.

The optimisation steps used in section 6.1.3 are repeated, starting the cladding (BGaP) layers, with the reasoning as per the previous discussion. With reference to figure 6.10 the same behaviour as for the case of the single quantum well structure is observed. The difference here is that the confinement factor levels out at $\sim 1.51\%$, with a layer thickness of 3 $\mu$m. Thus, once more the n-side BGaP cladding thickness has a negligible affect on the optical confinement factor but the p-side BGaP cladding continues to have a large impact as it determines to what degree the optical mode couples into the capping layer and how large a role the BGaAsP/air interface plays in guiding the mode. Since the physics behind the analysis of this plot is identical to that of figure 6.4 in section 6.1.3 it is not repeated for brevity.
Figure 6.10: Confinement factor vs p and n (black and blue) BGaP cladding thicknesses (varied individually). In each case the other layer thicknesses are set to their original values. The values for the n and p BGaP used in #16053 are marked by the dotted vertical lines (blue and black respectively).

The optical field and index profiles for #16053 after increasing the p-BGaP cladding thickness to 3 \( \mu m \) are shown below in figure 6.11. The mode is now symmetrical and well centered on the central quantum well; in fact the mis-alignment is only 21 nm. The leakage into the capping layer has also been eliminated. As such the confinement factor rose from 1.33% to 1.51%, with the difference in confinement factor of the individual wells dropping to fairly negligible levels. Once more the analysis of the impact of the improvements to the optical mode are identical to those relating to the single QW simulations. Also, as before the optimised cladding thickness will be larger than required once the barrier optimisation has been carried out.
Figure 6.11: Simulated optical field and refractive index for structure #16053 with a p-BGaP cladding thickness of 3 µm, \( \Gamma = 1.51\% \). QW/optical field mis-alignment = 21 nm. Top: Full device. Bottom: Zoomed to active region.

As before, the barrier thicknesses were then scanned from 0 to 1 µm in steps of 50 nm in order to further improve the confinement factor. It was found that the confinement factor was maximised for n and p side barrier thicknesses of 0.3 µm as was the case for the single QW structure. The confinement factor from 1.50% to 1.73% as can be seen in figure 6.15 as a result of the barrier optimisation, as can be seen from figure 6.15 below.
Figure 6.12: Confinement factor vs p-side barrier thickness for the standard #16053 n-barrier thickness of 0.15 µm (black) and optimised value of 0.3 µm (blue). The marked data are the standard #16053 barrier confinement factor (black) and optimised for both barrier thicknesses (blue).

Figure 6.13 indicates the relation of the optical mode to the confinement factor, by comparison to the previous iteration of the structure it can be seen that visually, the width of the mode when the field strength drops to zero has decreased from $\sim 4$ µm to $\sim 2.8$ µm. Better confining the mode has also led to the peak field strength increasing from $\sim 5$ to 5.8 (a.u.), a fractional increase of $\sim 14\%$ and with a QW/optical field mis-alignment of only 4 nm the increased peak intensity has a maximised effect on the confinement factor.
Visual inspection of the optical mode from figure 6.16 indicates that the optical field decays to zero approximately 1 \( \mu m \) before the capping layer. Thus a final batch of simulations was carried out, reducing the p-BGaP thickness from the first optimisation value of 3 \( \mu m \) down to the original 1.2 \( \mu m \) to determine the optimal value which minimises the device thickness without compromising on optical confinement factor. The result of this final optimisation step was that a p-BGaP thickness of 1.5 \( \mu m \) could be selected without a reduction in optical confinement factor. The result of the simulation of the final structure is shown below in figures 6.14. It can be seen that the optical mode remains centered on the active region with the mis-alignment remaining at 4 nm and the optical confinement factor remaining at its peak value of 1.73%. 

Figure 6.13: Simulated optical field and refractive index for structure \#16053 (MQW) with barrier thicknesses increased to 0.3 \( \mu m \) and the p-side BGaP cladding thickness increased to 3 \( \mu m \) in accordance with the findings of figures 6.12 and 6.15. \( \Gamma =1.73\% \). QW/optical field mis-alignment = 4 nm.
Figure 6.14: Simulated optical field and refractive index for structure #16053. Top: Standard structure, $\Gamma = 1.33\%$. QW/optical field mis-alignment = 94 nm. Bottom: Optimised with a P-BGaP cladding thickness of 1.5 $\mu$m and barrier thicknesses of 0.3 $\mu$m. $\Gamma = 1.73\%$. QW/optical field mis-alignment = 4 nm.

Once more the reduction in threshold gain, threshold carrier density and threshold current that would be expected from the increased optical confinement factor are approximated using equations 2.28, 2.32 and 2.21 respectively. The optical confinement factor increased from 1.33\% to 1.73, which yielded reduction of 23\%, 21\% and 27\% for threshold gain, carrier density and current respectively. The assumptions used in the single quantum well optimisation study apply here, giving the values an equally large uncertainty. As before though, even if the actual values are only a fraction of those approximated here the improvement in device performance (although less than in the SQW case) will be significant, particularly when to achieve this only reasonably minor adjustments to the structure are required.

### 6.1.5 Optical simulation conclusions

In this section the results of the passive simulation of GaNAsP based QW devices have been presented with the aim of maximising the optical confinement factor through the optimisation of cladding and barrier layer thicknesses. It was found that the optical confinement factor could be increased from 0.35\% to 0.6\% and 1.33\% to 1.73\% for the single and multiple quantum well structures, based on devices #16195 and #16053 respectively. The original and optimised structures of the single and multiple quantum well devices are shown below in figure 6.15, changed thicknesses are in bold.
In both cases the first change was to prevent the optical mode coupling into the capping layer by increasing the thickness of the p-side BGaP layer, increasing the value from 1200 nm to 1800 nm and 1500 nm for the SQW and MQW structures respectively. Secondly the barrier regions (BGaAsP) were varied in thickness in order to reduce the mis-alignment between the envelope peak and the quantum well(s). This was reduced from 110 nm to 3 nm for the SQW structure and from 94 nm to 4 nm for the MQW structure.

It has already been shown in previous work on this material system that increased confinement factor through the inclusion of additional quantum wells leads to an improvement in device performance [111] as well as the comparison of structures #16052 and #16053 in chapter 5. These simulations allow further increases in confinement factor to be realised without the need for major structural changes, only slight alterations to the barrier and cladding layer thicknesses. The ratio between the unoptimised single and multiple QW structures was initially $\sim 3.8$ but post optimisation the value falls to $\sim 2.9$ indicating that increasing the number of QWs increases the confinement factor by the expected ratio once the mis-alignment is eliminated.

This also highlights that the optimisation of the two structures, which was fairly similar in both cases, has a significantly greater impact on the single QW structure.

Approximating the reduction in threshold gain, carrier density and current for both structures yielded
values of 42%, 35% and 41% for the SQW device and 23%, 21% and 27% for the MQW structure. The assumptions required to calculate these parameters make the uncertainty on the absolute values somewhat large. However, the values are large enough that even if in reality they are only a fraction of these the improvements will remain significant, particularly because minimal changes are required to achieve them.

Thus, the structural optimisations described in this work represents a quick and efficient method of improving device performance through increased modal gain and by extension reduced threshold current density. A reduced threshold current density should then reduce the fractional contribution of highly carrier density dependent loss mechanisms such as carrier leakage, which has previously been found to dominate these devices and is discussed in chapter 5. Movement of the optical field away from the p-doped regions will also reduce optical losses, as significant photon density in an area of p-doping will lead to free carrier absorption.

Growth of the optimised structures described in this section would allow verification of the results of this investigation through experimental device characterization experiments. The optical confinement factor may also be increased through enlarging the index steps between the barrier and cladding regions of the structure. An increase in the index contrast could be achieved through reducing the Boron content in the cladding and/or increasing the Boron or Phosphorous content in the barriers.

The problem with this approach, however is an associated reduction in carrier confinement as well as potential issues with strain and thermal expansion coefficient management. The improvement of optical confinement would therefore have to be balanced with the detrimental effects on strain and carrier confinement. It was shown in section 6.1.3 that due to the inherently similar refractive indices of GaAsP and GaP that large changes in composition are required to achieve an increase in optical confinement factor of any practical significance.

It is also possible that a superlattice of GaP/BGaP in the cladding adjacent to the barriers could improve confinement with a lesser effect on strain and carrier confinement. This however could impact carrier mobility, especially in the case of holes, which may become trapped. The LaserMod software package is capable is simulating the full operation of a laser structure, including the effects of defects and strain, making these possibilities potential avenues of future investigation. However, since a full material file is required for this kind of simulation, including a list of over 100 individual material parameters this is currently not a possibility since many of these parameters remain unknown for the novel materials GaNAsP, BGaP and BGaAsP. This remains the subject of future work.
6.2 GaNAsP/BGaP/Si wafer processing procedure development and device characterisation

6.2.1 Identification of the processing limitations of post structure #16195 devices through optical and scanning electron microscopy

It was known from time spent working with the growers in Marburg that the focus of the group was on producing high crystalline quality wafers of the most efficient and optimised structure. The processing of the wafer material into bars of broad area lasers is an area that required further study. With this in mind, when the post #16195 devices were found to be consistently non-functional they were investigated using SEM to determine if there were any visual indicators of cause for this.

It had already been observed with the unaided eye and optical microscope that when devices failed it was often as a result of the contacts (both p and n type) “burning” at both high and low injection currents. Since past devices required several amps of current to achieve threshold, it was frequent that in electroluminescence investigation devices would be destroyed in this manner. An example of such contact burning is shown below in figure 6.16. Sometimes the current probe could be moved to a different region and the device would continue to function, but often a burnt contact would mean the device became Ohmic when biased thereafter.

The burning of the contacts suggests an issue with the metalisation of the devices. An inhomogeneous deposition of metals and/or the presence of contamination in some form, which would cause hot spots and burning as a result of current flow when biasing are possibilities. Certainly the resistance must be large in those devices where contact burning took place for a sufficient quantity of energy to have been absorbed in the contacts for burning to take place. In addition when the contacts did not burn the IV characteristics were found to be consistently poor, with a large and poorly defined turn on voltage. Example IVs are presented in appendix D.
Figure 6.16: The contacts of two structure #25486 devices after attempting electroluminescence with a 0.5% duty cycle and a 25 \( \mu m \) probe to contact. Damaged, burnt sections can be seen at multiple points which were caused by injection currents of less than 100 mA, far less than the currents exceeding an Amp applied to older devices.

Using the SEM each bar of devices was investigated visually in the hope of identifying limitations related to the processing. The first type of abnormality found was a range of crystalline deposits across a range of devices and structures. Figure 6.17 shows two such examples from structures #25487 and #25488. It should be noted here that prior to characterisation as well as this SEM investigation the devices were cleaned using acetone, isopropyl alcohol and deionised water to remove any contamination. This cleaning procedure matches that used by the growers and is common for semiconductor devices, it would therefore not be expected to introduce additional contamination.

In the left hand image some deposits with a size of \( \sim 20 \mu m \) are visible across both the surface of the p-type gold contact (lighter material) as well as the semiconductor surfaces. In the right hand image another crystalline deposit can be seen to have spread over a larger area of the device but with a finer grain of \( \sim 1 \mu m \). Having cleaned the devices with acetone, IPA and deionised water it is unexpected that crystalline deposits would be found on the device surface, since these solvents are specifically chosen to remove the majority of chemicals that the devices could reasonably be expected to come into contact with.

With no immediately obvious identity for the deposits it was decided to use energy dispersive X-ray spectroscopy (EDX) to characterise the chemical composition of the surface in the region of the crystalline material. Strangely, other than the semiconductor constituents expected to be present in these devices strong signals from potassium and chlorine were detected in the region of these deposits, but not in other areas of the device surface. A possible origin of the deposit is the reaction of potassium hydroxide (the developer for the AX100 photoresist used at Marburg) and the etchant, which is 30% HCl.

Potassium chloride is very soluble in water and so it is unusual that this would be present on the devices...
after multiple stages of cleaning with deionised water, among other solvents following etching. This then seems unlikely, another possible point in the processing that could produce potassium chloride is just before contact deposition. At this stage the surface is etched with 30\% concentration HCl to ensure it is free of oxidisation before metalisation, if the developer were not completely removed then it could potentially react with HCl at this stage. Again though, even the etchant itself is 70\% water and would be expected to remove the potassium chloride.

In addition the deposits are found both on the metal contacts as well as the un-metalised semiconductor surfaces, which would have been isolated from the etchant, again making this source seem unlikely. Therefore it is difficult to say where precisely the deposits originate from, but it is likely at some stage during processing rather than after the devices were completed since they are unlikely to have come into contact with the required chemical constituents for potassium chloride formation.

Further, the deposit was found both on devices which were sent to the University of Surrey as well as those which were simply cleaved from the wafers and stored at Marburg. There were no other strong signals during the EDX investigation that would indicate the presence of a material that was preventing the potassium chloride from dissolving and so it is unclear how it could have remained on the device surface. Without being able to say more on the reason for which the potassium chloride was not dissolved from the surface the main conclusion that can be drawn is that regardless of the exact origin, the deposits should not be there, and are an indication of a short coming in the processing methodology.

Figure 6.17: SEM images of two devices of structure #25487 indicating a crystalline build up of both large (10 \(\mu\text{m}\), left image) and fine (<1 \(\mu\text{m}\), right image).

The second category of observation made under SEM was poor definition of the etches which form the mesa structure. Figure 6.18 shows two examples of problems common to the devices received post structure #16195. In the left image the mesa can be seen to extend over into the region metalised for n-type contacts. This is likely a problem with the lithography stage, which defines the areas protected from the etchant by
If there is a gap between the mask and photoresist, the photoresist thickness is inhomogeneous, contaminants are present, the sample is over exposed or over developed additional, unwanted etching may take place.

Conversely insufficient exposure due to too low a UV light intensity, too short an exposure time or insufficient developing could cause incomplete dissolving of the photoresist and etching to a less than desired extent. The right hand image shows a general inhomogeneity to the mesa sides, with regions of differing etch depth that could be caused by either incomplete and poorly defined exposure of the photoresist or an inconsistent etch rate. The side wall parts on the right hand image with a granular appearance were at first thought to be another deposit of some sort, but EDX found no materials other than the semiconductor constituents of the device indicating that this is not the case. In all images shown here the mesa appears to have two parts, a main section and a slightly lower section along both sides with a width of approximately 10 \( \mu m \). This is in fact not an intentional feature but an area of excess lateral etching.

![Figure 6.18: SEM images of two devices of structure #25487 indicating a poorly formed mesa structure.](image)

In the left image we see that the mesa extends over into the region set aside for the n-type contact and as a result is partially metalised. On the right a device can be seen for which the mesa sides are rough and not completely formed.

Since the etchant will attack all material it comes into contact with it is expected that some degree of lateral etching will take place as the side walls are exposed. Lateral etching leads to the feature becoming thinner at the top than the bottom, flaring out according to the amount of time the side walls have been exposed to the etchant. This gives the overall shape of the image shown in figure 6.19, but assumes that all material etch at a single rate. The differing chemistry of the various materials that make up the structure cause differing etch rates for each of the layers.

For the aqua regia based solution used at Marburg the etch rates for GaP and GaAs are approximately 2 \( \mu m \) min\(^{-1}\) and 6.2 \( \mu m \) min\(^{-1}\). If the rate is assumed to vary according to linear interpolation and the boron and nitrogen content assumed to have negligible impact, then the etch rates for GaNAsP and BGaAsP
would be expected to be approximately 240% and 30% faster than that of BGaP, which makes up the bulk of the structure. If these regions of faster lateral etching are taken into account then the overall etch pattern will look like the cross section depicted in figure 6.19 (not to scale).

Thus, when viewed from above the lasers will appear to have the additional step along the mesa edges that is visible in all SEM images as the BGaAsP capping is laterally etched at a greater rate than the BGaP below. The more important impact however is the lateral etching of the active region, since this will reduce the gain producing volume that is available to be pumped and increase the current density within the central region that remains. The result is that the threshold current densities may, in fact, be even larger than is thought currently from electroluminescence studies.

In addition etching of the quantum well will create defect states on the surface which may contribute to the non-radiative recombination which dominates these devices. It has been reported by Marburg that in some cases undercutting of the quantum well has been observed to be as much as 70%. In addition, this undercutting reduces the mechanical strength of the structure and may lead to the sides of the device breaking off since they are unsupported below. The undercutting may be exacerbated beyond the increased etch rate of the material relative to the BGaP by the fact that the region is so narrow, inhibiting the etchant from being completely flushed out with water post etch. Since the aggressive aquaregia requires only around 30 s to define the structure, even a second more exposure could cause considerable damage to the quantum well.

![Figure 6.19: A depiction of the undercutting taking place when the GaNAsP based lasers are etched using Aquaregia solution.](image)

The final observation made by SEM was the inhomogeneity and inconsistency of the gold contacts. Figure 6.20 displays two examples of devices with rough, granular contacts. This figure highlights that the deposition of metals is not consistent across the wafer, since the devices shown in figures 6.17 and 6.18 although not having perfect contacts are considerably better than these. Furthermore, for the contacts to be this rough either the evaporator parameters must be non-optimal or a source of contamination is being introduced into the system at some stage in the deposition.

The procedure, equipment, evaporator parameters (coil voltage, deposition time etc) and metals used to deposit the contacts on these devices is identical to that used for the #16195 and prior structures as well as those produced as a part of separate projects, for which the contacts have not been found to exhibit these
issues. Thus contamination of the evaporator or metals around the period of time these devices were grown would seem like the most logical cause for these observations.

It is possible that the metalisation issues could be linked to the deposits of KCl observed in figure 6.17. If there were KCl on the surface of the sample when metalisation took place this would lead to poor adhesion and defect containing, insulating regions that may cause hot spots and the observed contact burning. The metals chosen for the p-type and n-type contacts were a 20 nm Cr wetting layer followed by 500 nm of Au, and a 20 nm Cr wetting layer followed by 400 nm of AuGe respectively. The least ideal contacts are always p-type suggesting that it is this contact is more problematic than the AuGe n-type contact. It should be noted that there is no post metalisation annealing step for these devices, which could potentially improve the contacts to some degree by homogenising the lattice.

Figure 6.20: SEM images of two devices of structure #25487 indicating the poor quality of the contacts (left) and a device which exhibits all processing shortcomings of the devices.

The right hand image of figure 6.20 is of particular interest since it contains all of the issues discussed above as well as emphasising the need to optimise the device processing methodology. On top of the above issues, one problem inherent to the material system is the rough side wall. Cleaving these devices is challenging since Si is well known to be brittle[120]. The main limitation of these devices is the strong non-radiative recombination resulting from the physical properties of the materials combined to form these devices, as is discussed in chapter 5. However, to be able to experimentally characterise future devices and in doing so identify steps to be taken to reduce the non-radiative recombination, the processing of the devices must be improved sufficiently to not interfere with the characterisation process or results.

6.2.2 Wafer-processing optimisation investigation

Section 6.2.1 served identified several limitations of the post #16195 GaNAsP based devices that have been attributed to un-optimised device processing. With this in mind it was decided that an investigation into
optimising the processing methods for GaNAsP based lasers should be carried out at the University of Surrey, where the clean room facilities are more suited to carrying out this type of work. As a reference the processing process flow at the University of Marburg was as follows:

1. Cleaning: Boil in acetone, then deionised water, then isopropyl alcohol (IPA) then deionised water once more for 5 minutes each. Dry with nitrogen gun followed by 1 minute on the hot plate at 100 °C.

2. Photolithography: AZ4562 thick-film photoresist is pipetted onto the wafer surface and spun at 4000 RPM, depositing \( \sim 8 \mu m \) of resist, then left to de-gas for 15 minutes followed by a bake of 1 minute per \( \mu m \) of resist at 100 °C. The wafer is placed in the mask aligner and the mask brought into contact with an exposure time of 50 s. A 1:4 mixture of the AZ400K developer and deionised water is used to remove the unexposed photoresist for \( \sim 2 \) minutes.

3. Etching: Aquaregia (HCl:HNO\(_3\)) is prepared and the wafer dipped into the etchant and observed. As the mesas become visible and a colour change is observed indicating that Si has been exposed the sample is removed. A deionised water kill-bath is used to end the reaction. The photoresist is then removed by submersion in the AZ100 remover solution and washed with deionised water.

4. Metalisation: Following a repeat of the photolithography stage and immediately before placing in the evaporator the sample surface is etched using 30% concentration HCl and 3% concentration HF for the III-V and Si surfaces respectively. With the shutter closed the coil current is raised to 42 A and once the chosen metal is observed to have melted and a small signal detected the shutter is opened to expose the wafer to the metal vapour. Cr is used as a wetting layer for both the p and n contacts and is deposited at a rate of 0.2 \( nms^{-1} \). The next metal is selected (Au for p type contacts and AuGe for n type contacts) and the deposition repeated at a rate of 0.6 \( nms^{-1} \). Different evaporators are used for p and n type metals to avoid cross contamination. In total 20 nm of Cr is deposited for both contacts, on top of which 500 nm and 400 nm of Au and AuGe are deposited for the p and n-type contacts respectively.

These steps are used as a starting point from which an optimised processing methodology can be derived after suitable investigation of each aspect. The investigation is broken down into lithography, etching, metalisation and the dicing of wafers into individual devices. Following the processing the devices were then characterised to determine the degree of success of the optimisation process. Cleaning using acetone, IPA and deionised water is a standard process and as such is used without significant changes.

Before the investigation into processing methodology could be started masks were first designed and manufactured with the intention of being able to fabricate a selection of different devices even if a piece of wafer as small as \( \sim 1\times1 \) cm is used. This was necessary since the supply of wafer material is limited (some structures are grown on only 1/4 of 2 inch wafers, only a fraction of which can be expected to be sent to the University of Surrey. The chosen features were \( 100 \mu m \) and \( 200 \mu m \) mesa width stripes for producing broad area lasers, matching those produced at Marburg, as well as \( 50 \mu m , 75 \mu m \) and \( 100 \mu m \) circular LED
structures (see figure 6.21 for an example) enabling a more diverse range of experimental characterisation experiments.

### 6.2.2.1 Lithography

One potential issue with the lithography process described above is the use of a very thick (8 $\mu m$) and viscous photoresist, which leads in turn to a comparatively large edge bead when compared to thinner photoresists. The edge bead forms at the perimeter of the sample after spinning as the result of the centrifugal force that is used to spread the resist across the wafer. On a large sample the proportion of the surface with an edge bead is small, but since only small (at the very most 2 inch wafers, but generally only a few square cm) samples are processed in project the edge bead surface area proportion is significant.

The presence of the edge bead has two main detrimental affects. Firstly, since the resist is thicker the region beneath the edge bead may not develop sufficiently and thus cause the photoresist to not accurately match the mask pattern. Secondly, when exposure takes place the mask must be brought into contact with the photoresist surface, if there is an edge bead present this will act as a spacer, preventing good contact across the sample. If there is a gap between the mask and surface then photons of non-normal incidence may leak underneath the mask slightly leading to poor resolution of the photoresist mask.

The height of the edge bead can be reduced by using a thinner, less viscous photoresist. However, the reason for using the thick resist in this case is to offer enough resilience to the aggressive etchant (aquaregia) used to define the mesas in these devices. The aggressive nature of the etchant and particularly its tendency to laterally etch the structure due to the differing etch rates of the BGaP, BGaAsP and GaNAsP is another aspect of the processing to be investigated. Thus it was decided to optimise the lithography by changing the photoresist since the etchant was likely to be changed also.

The photoresist chosen as a potential alternative was AZ5214E, consisting of novolak resin with naphthoquinone diazide as the photoactive compound. This resist is commonly used within the ATI when processing a variety of material systems and has been found to produce good results. When spun at 4000 RPM this resist gives a thickness of approximately 1.4 $\mu m$ and can be used as either a positive or negative resist by adding an additional high temperature bake to the lithography process. Further details of AZ5214E can be found in the data sheet [1]. Using the data sheet and advice of those with experience in the use AZ5214E the following lithographic procedure was developed:

1. Cleaning with Acetone, followed by IPA and finally deionised water followed by drying with first the dry nitrogen gun and then a 1 minute bake at 100 °C on the hot plate.

2. The sample is placed in the spinner and the vacuum activated to secure it in place, the primer hexamethyl distillazine is flooded on and spun at 4000 RPM with an acceleration of 11440 $RPMs^{-1}$ for approximately 12 seconds at which point the colour change resulting from thin film diffraction is observed to stop. AZ5214E is flooded on and the sample is spun once more with the same settings but for 30 seconds. After removal from the spinner a soft bake at 95 °C for 1 minute ensures the evaporation
of the solvent carrier.

3. The sample is then placed on the vacuum stage of a Suss MA1006 mask aligner and the chosen mask (three were available, one for mesa definition, one for p-type metal and one for n-type metal) fitted and aligned with the sample. The exposure parameters were a UV lamp intensity of $7 \text{mWcm}^{-2}$, hard contact between sample and mask and an exposure time of 5.5 seconds.

4. The exposed sample was then developed using a 1:4 mixture of AZ351B developer and deionised water by immersion for 35 seconds and the reaction halted using a deionised water kill-bath.

The parameters such as developer time, exposure time, contact type and contact type were gradually optimised by visual inspection of the developed resist under optical microscope. This process was quick and effective since the resist could simply be removed and the process repeated with varied parameters until acceptable reproduction of the mask pattern was achieved. Of particular importance were the time exposed to the developer and the UV light exposure time.

Too much UV or developer exposure would lead to the removal of too large a quantity of resist and therefore smaller feature sizes than desired. Too little UV or developer exposure can lead to features that are either larger than desired or a photoresist residue between features.

In both cases the features would generally poorly resolution with a loss of fine detail. An example of the resolution achieved using the optimised lithography methodology is shown below in figure 6.21. Here a Surrey stag identifier mark is shown alongside a 50 $\mu$m diameter LED p-type contact defined in photoresist. The good resolution and accurate reproduction of the stag, where features in the antlers are less than a micron across indicates a well optimised lithography process.

![Figure 6.21: The accurate reproduction of a stag, University of Surrey identification mark is shown alongside a 50 $\mu$m diameter LED p-type contact defined in photoresist.](image)

### 6.2.2.2 Etchant investigation

As discussed in section 6.2.1 the choice and application of etchant is an area of particular concern since there is evidence of lateral etching of the capping layer. This is likely to have
also taken place to some extent in the active region, since the barriers and quantum well are equally and even more reactive that the capping layer respectively. This lateral etching is undesirable and so a number of alternative etchants were investigated in order to determine which was most suitable.

Many etchant choices exist for the variety of materials commonly used to form semiconductor devices, summaries for which can be found in the following literature [87][36]. Of the many potential etchants six were short listed for this investigation and are listed in figure 6.22. The two main materials from which these lasers are composed are BGaAsP (barriers and capping) and BGaP (cladding), the boron content is fairly low (\(\sim 5.6\%\)) and so etchants were chosen that have been shown to work well with GaAsP or GaP.

The first three etchants from the table in figure 6.22 were chosen from [87] to represent a range of etchant chemicals and etch rates that are known to produce good results for GaAsP. The fourth etchant is described in [74] as an effective etchant that leaves a smooth surface with minimal lateral etching under the correct conditions. The fifth and sixth etchants were both taken from [43], where a comparison of etchant 5 (a less aggressive version of the etchant used at Marburg) is made to etchant 6; suggesting that the addition of acetic acid to the hydrochloric and nitric acid gives a more gradual etch leading to a superior surface quality.

The presence of arsenic in the BGaAsP (and particularly GaNAsP) unlike the BGaP means that the chemistries of these materials are relatively different and so the etchants were tested on both materials independently to determine if they were effective in both cases. If it is found that no single etchant is suitable for use on both materials then the etching process will have to become a more complex multiple stage processes, which is to be avoided where possible.

<table>
<thead>
<tr>
<th>Target material</th>
<th>Composition</th>
<th>Mixture ratio</th>
<th>Etch rate from literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAsP @ 300K</td>
<td>(\text{H}_2\text{O}_2:\text{HCl}:\text{H}_2\text{O})</td>
<td>5:2:2</td>
<td>1 (\mu\text{m min}^{-1})</td>
</tr>
<tr>
<td>GaAsP @ 300K</td>
<td>(\text{H}_3\text{PO}_4:\text{H}_2\text{O}_2:\text{H}_2\text{O})</td>
<td>6:3:100</td>
<td>0.1 (\mu\text{m min}^{-1})</td>
</tr>
<tr>
<td>GaAsP @ 300K</td>
<td>(\text{HNO}_3:\text{H}_2\text{O})</td>
<td>1:2</td>
<td>0.8 (\mu\text{m min}^{-1})</td>
</tr>
<tr>
<td>GaAsP @ 300K</td>
<td>(\text{HNO}_3:\text{H}_2\text{SO}_4:\text{H}_2\text{O})</td>
<td>3:1:8</td>
<td>0.66 (\mu\text{m min}^{-1})</td>
</tr>
<tr>
<td>GaP @ 330K</td>
<td>(\text{HCl}:\text{HNO}_3:\text{H}_2\text{O})</td>
<td>2:1:2</td>
<td>1 (\mu\text{m min}^{-1})</td>
</tr>
<tr>
<td>GaP @ 330K</td>
<td>(\text{HCl}:\text{HNO}_3:\text{CH}_3\text{COOH})</td>
<td>1:1:1</td>
<td>0.25 (\mu\text{m min}^{-1})</td>
</tr>
</tbody>
</table>

Figure 6.22: The etchants and expected etch rates for their target materials investigated for use with the GaNAsP material system.

The first etchant investigation was carried out on a test structure consisting of a thick layer of BGaAsP (\(\sim 1.4 \mu\text{m}\)) on a GaP nucleation layer, epitaxially deposited on Si. Lithography was carried out to define the mesa pattern onto the surface of \(\sim 1.5 \text{ cm}^2\) wafer samples and subsequently cleaved into six pieces to minimise material consumption. These pieces were then submersed into the etchants for varying periods of time and an AS-200 AlphaStep surface profiler used to measure the depth of the features.

These measurements were taken at 3 different positions on the surface of the sample, then averaged and compared to those taken prior to exposure to the etchant. If no difference was measured the samples were re-exposed to the etchant for extended periods to determine if no reaction took place or it was simply slow.
The results of these initial tests are plotted in figure 6.23. Of the six etchants tested only three were found to have any effect on the sample (four, five and six).

The two etchants designed for use with GaP were found to react with the BGaAsP slightly faster than would be expected for GaP from the literature. This was expected since the processing carried out at Marburg used a similar etchant and exhibited a more rapid etch rate for BGaAsP and GaAsNP. Surprisingly, only one of the four etchants tested here that are commonly used with GaAsP was able to etch the BGaAsP sample, and the one that did exhibited an etch rate less than a third of that expected from the literature.

Two possible explanations for this were considered. The samples were fairly old and had been exposed to air for over a year, meaning that they may have had an excessively thick oxide layer on the surface inhibiting etching. This hypothesis was tested by carrying out an extended surface oxide etch using 30% concentration HCl, as is commonly carried out for both GaAs and GaAsP since these oxides are reactive with HCl. Incidentally this is the use of the HCl in the etchants that contain this acid.

Hydrogen peroxide and nitric acid are both oxidisers and so the reaction of the HCl containing etchants is to first oxidise the semiconductor surface and then dissolve the oxide. Application of concentrated HCl for 30 minutes ensured that if any thick oxide layer were present it would be removed before successive etchant investigations were to take place. Further etches after the application of the HCl yielded identical results, indicating that the presence of a surface oxide was unlikely the cause of the low reactivity of the samples in the chosen etchants.

The second hypothesis was that the boron content may have had a non negligible effect on the chemistry of the material and inhibited the reaction of the etchants with the As and P. This was somewhat more difficult to test and unsupported by literature since the boron containing materials have not been the subject of a great amount of detailed study. As such it was decided to continue the investigation by following up with the etch test on a BGaP sample to determine the effectiveness of the etchants on this material and see if any further evidence could be found to support this hypothesis.
Figure 6.23: The initial etch tests for a range of wet etchants applied to BGaAsP test material, used to determine which are suitable for further investigation.

The tests on the BGaAsP sample were repeated on a similar BGaP sample (thickness \( \sim 1.4 \, \mu m \)) and the results plotted in figure 6.24. Similar results to those on BGAsP are found, but with the exception of the sulphuric acid based etchant (number four in the table) having no effect on the sample. This etchant is similar to number five, with the difference being the quantity of water and the exchange of hydrochloric for sulphuric acid. Thus it would seem that the sulphuric acid is not suitably reactive with the oxide produced by the reaction of nitric acid and BGaP, leading to no etching.

The two remaining etchants were found to have rates comparable to those found in the literature, 5 slightly less and 6 slightly more. Since these were rough initial tests with only a few data points these numbers are unlikely to be exact and so the differences are to be expected. Since the sulphuric acid containing etchant had no effect on the BGaP no further information as to why its rate was so much lower than expected for BGaAsP is derived from this test. However, since this etchant is unreactive with the bulk of the laser structure it is now discounted from the investigation unless a two part etch is deemed to be necessary.
Figure 6.24: The initial etch tests for a range of wet etchants applied to BGaP test material, used to determine which are suitable for further investigation.

In order to choose between the two remaining etchants the surface quality was inspected using a SEM to determine if any discernible difference could be found which would indicate one of the two to be the superior. Figure 6.25 shows the surface images on the etched regions of samples exposed to etchant 6 (left) and 5 (right) with a scale of 50 µm. It is immediately evident that the etchant containing acetic acid in addition to hydrochloric and nitric acid leads to a far smoother surface, free of defects and pitting.

Surface quality is an important factor when considering metalisation since voids and defects in the surface will lead to poor contacts that may have a higher than optimal resistance or even develop hot spots and a non uniform current spreading throughout the device. In addition, rough and defect dense mesa side walls will lead to reduced optical confinement of the lateral mode and may increase monomolecular recombination. Given that both etchants effectively removed material from the BGaAsP and BGaP samples as well as the fact that the etch rate difference between the two materials with the acetic acid etchant is lower, etchant 6 is the superior choice.
The pitting could have resulted from insufficiencies in the preparation or application of the etchant and so several repeats were carried out including a range of exposure times. The results were much the same in each case and indicated that this effect may well be inherent to the combination of etchant and target material. Within the literature from which these two etchants were taken ([43]) figure 6.26 was presented, showing similar results for aquaregia and etchant 6 when applied to GaP. The explanation relates to the somewhat unusual nature of etchants similar in nature to aquaregia, which contain a mixture of hydrochloric and nitric acid. When these two acids are mixed the following reaction takes place between the acids themselves:

\[ 3HCl + HNO_3 \rightarrow NOCl(g) + 2Cl(g) + 2H_2O(l) \]  

\[ 2NOCl(g) \rightarrow 2NO(g) + Cl_2(g) \]

\[ 2NO(g) + O_2(g) \rightarrow 2NO_2(g) \]

In fresh aqua regia the nitric and hydrochloric acids react with the semiconductor according to the oxidise (with \( HNO_3 \)), dissolve oxide (HCl) pattern of material removal common to many etchants. However, the acids react with themselves at the same time forming the products NOCl (nitrosyl chloride, a yellow gas which dissolves to give the characteristic colour of aqua regia) and chlorine. The chlorine gas will slowly rise through the solution and escape into the atmosphere (causing the observed fuming of aqua regia). However, if it is formed at the surface of the semiconductor then it may also react with the As or P.

Chlorine, being only second in elemental reactivity to Flourine will cause an increase in local etch rate as bubbles are formed and therefore cause the inconsistent pitted surface of the sample. Pits in turn cause an even greater local etch rate as the surface area exposed to the etchant is increased, exacerbating the
problem. The rates of the reactions shown in 6.8 are linked to the rate of change of reactant and product concentration. The initial decomposition of the acids will be rapid, but slow to an equilibrium rate once the solution is saturated with NOCl. At this stage the rate of decomposition will be limited by the secondary decomposition reaction rate of NOCl into NO and $Cl_2$.

Therefore to minimise pitting and achieve a more consistent etch rate the solution must be left to stabilise until this equilibrium is achieved. The reactivity will slowly decrease from this point as the acids are decomposed and so the solution must be used as soon as it is ready and not kept to be used at a later time. The useful lifetime may however be extended somewhat by sealing the solution in an air tight container to maximise the concentration of the decomposition products and thus minimise the breakdown rate.

The addition of acetic acid as a solvent for the hydrochloric and nitric acids serves to produce a less aggressive etch by reducing the concentration of etchant without inhibiting the breakdown of the acids. If the solution is diluted with water, although this reduces the acid concentration it also increases the concentration on the right hand side of the reaction on the top line of 6.8. This therefore causes a slower initial breakdown of the acids but in addition since chlorine gas reacts with water it will cause an increase in the re-formation of HCl. Thus, reducing the acid concentration with the addition of acetic acid reduces the overall etch rate, the chlorine gas concentration and formation rate, and the difference the formation of pits makes to local etch rate.

![Figure 6.26: Figure taken from][43] pitting as the result of a diluted aqua regia etch compared to the smooth surface of the same etchant with the addition of acetic acid.[159]

With a less aggressive etchant chosen tests of etch rate through the structure containing all three materials were undertaken in order to calibrate the etchant. The samples were prepared as previously and two types of etch test considered in order to account for all possibilities when proceeding to the actual device processing runs. Firstly, the etch rate was investigated for both stabilised and fresh solutions in order to determine how much of a difference this would make and if there were any major differences between the two approaches.

As can be seen from figure 6.27 the fresh solution exhibits an etch rate of 900 nm $min^{-1}$ in contrast to the 306 nm $min^{-1}$ observed for the solution when stabilised for 30 minutes. Interestingly the fresh solution
appears to behave almost identically to etchant 5 in terms of etch rate. Since initially the acid concentration will be much greater this is somewhat to be expected. Pitting was also observed under microscope indicating that the rate of chlorine generation was also comparatively large.

For this reason the fresh solution was discounted as a potential etchant also, leaving only the slower, stabilised solution. The stabilised solution, with an etch rate of 306 nm $min^{-1}$ was found to produce a smooth surface, free of pitting and allowed better control of etch depth since timing was not required to be as exact as for a faster etchant ($\sim 5 \text{nm}s^{-1}$ compared to $\sim 15 \text{nm}s^{-1}$). In both cases the etch rate as a function of time was near constant, making calculating the etch time required to achieve a desired feature depth simple and precise.

![Figure 6.27](image)

Figure 6.27: The etch rates for fresh (black) and stabilised for 30 minutes (red) HCl:HNO$_3$:CH$_3$COOH etchant on a piece of wafer with a complete laser structure.

The effect of consecutive etches was also tested to determine if subsequent etches could be applied to a sample who’s features may not have been properly defined. Figure 6.28 shows that multiple etches do not produce the same etch depth as single step etches of the same total immersion time. The reaction rate is likely not constant at the onset as a result of the two part nature of the reaction and the fact that the samples behaved somewhat hydrophillically, preventing the solution reaching the surface immediately. As a result the use of multiple exposures of single samples for device fabrication was deemed possible but somewhat unreliable if only a few 10s of nms were required to be removed.
The ideal target etch depth for the processing of lasers was determined to be 2.5 µm for the single quantum well structure received from Marburg, (identical to #16195, see figure 5.6) as this placed the n-type contact 400 nm into the highly doped ($10^{19}$ cm$^{-3}$) n-type BGaP layer. This differs from the current approach, where the etch is stopped when a visible colour change signifying the exposure of Si is observed. This approach was likely chosen to fit the highly aggressive etchant that would have made achieving precise etch depths challenging. By stopping at the Si substrate uniformity between batches is achieved since Si is unreactive with Aquaregia.

Using this approach the current must be injected through the defect dense interface between the III-V and Si regions. In doing so monomolecular recombination and scattering of carriers will take place which will reduce the carrier density at the QW, requiring stronger pumping to achieve threshold as well as dissipating significant heat at the interface. In addition if the defects are not uniformly distributed across the interface the current may preferentially flow through low resistance sections causing large current densities at these points which may cause damage to the devices. The studies presented in chapter 5 as well as previous studies at the University of Surrey on GaNASP based lasers have universally found defect related recombination to be the dominant source of loss. Thus moving the n-contact to the III-V region may have considerable benefits. Using a timed etch based on figures 6.27 and 6.28 the desired etch depth could be reliably achieved to within ~50 nm, well within the margin for error of the layer thickness.

A photo taken through an optical microscope is shown in figure 6.29 (left) of a representative metalised sample exhibiting the mesa definition and surface quality achieved using the process developed here. For comparative purposes a device from the batch of structure #16195 is also shown (right) since this is the best
performing of the GaNAsP based lasers. The mesas of the left hand image do not exhibit the step along both sides resulting from lateral etching of the capping layer that the right hand image does.

This alone highlights several important points about the etching of these devices. The lack of lateral etching of the capping shows that the comparability of the etch rates for BGaP and BGaAsP resulting from the change of etchant has lead to an improvement in the accuracy of the etching process. As a result, since the capping layer (the part exposed longest to the etchant) has not undergone noticeable lateral etching the BGaAsP barriers and GaNAsP QW should also not have been undercut. The potential undercutting of the active region was one of the main concerns related to the previous processing procedure since maximising crystalline quality (particularly for a material system known to be dominated by defects) and reproducing the structure exactly as designed is paramount to maximising performance or even achieving lasing at all.

Figure 6.29: Final etch quality, shown here for a fully metalised sample (left) alongside a mesa and n-contact from a bar of structure #16195.

6.2.2.3 Metalisation

The main goals of the metalisation study were to address the burning of contacts, the rough, inhomogeneous p-metal and to place the n-type contact within the highly doped n-BGaP layer rather than the Si and through these potentially improve the IV characteristics. The metals used to form contacts in this investigation were chosen based on literature [13] as well as advice of staff members at the University of Surrey who frequently metalised GaP and GaAs based semiconductor devices. The p and n-type contacts were investigated individually since they will use different metals, dopants and be deposited upon different materials.

Before all metalisation runs the samples underwent an etch in 10% concentration HCl/$H_2O$ in order to remove any native oxide build up formed as a result of exposure to atmospheric oxygen and were immediately placed in the evaporator and a vacuum applied to prevent the re-formation of surface oxide.

The p-type contact featured both the rough surface and contact burning, making it the more crucial to improve upon. Since the p-type dopant for the BGaAsP and BGaP was zinc this was chosen to be deposited
first as a dopant to be diffused into the BGaAsP capping and reduce the Schottkey barrier height at the semiconductor/metal interface. The first attempt to deposit p-metal was carried out using an Edwards E306A metals evaporator (the same system used at Marburg, the operation of which is detailed in section 3.5.3.1 of chapter 3) into which two titanium boats were loaded, containing zinc and pure gold.

Initially only small amounts of gold were deposited whilst making test contacts in order to minimise the consumption of a limited and expensive resource. As such 500 Å of Zn and 1200 Å of Au were deposited as reported by the oscillating crystal thickness sensor within the evaporator. Lift-off was carried out by leaving the sample to soak in acetone over night, dissolving the photoresist and removing the excess metal. However, on returning to the sample and rinsing in IPA and deionised water it was found that the metal had failed to adhere to the sample and was removed entirely.

Two potential hypotheses for the failure of the metals to adhere were considered, firstly that the surface oxide may not have been completely removed and secondly that the lack of thickness may have led to the metals peeling off as the photoresist was removed. In order to deal with these it was decided to increase the thickness of gold deposited, since using a little more was still more efficient that wasting an entire sample and the native oxide etchant was increased in concentration from 10% to 30% HCl.

Two more attempts were made to deposit Zn doped gold contacts on the mesa with good lithography achieved on both occasions. On the first attempt the Zn was observed not to melt but instead appear to start to sublime, before popping out of the boat before full deposition, a thickness of only \( \sim 90 \, \text{Å} \) was recorded by the sensor. The violent movement of the Zn pellet was attributed to the tendency of the material to suddenly sublime rather than melting and evaporating in an easily controllable manner. Gold was deposited to a thickness of 2000 Å but once again on liftoff no adhesion was achieved.

The second additional attempt was abandoned following the Zn pellet completely subliming in a flash, causing a cloud of hot Zn vapour to fill the evaporator. On removal the sample was found to be coated in what appeared to be zinc oxide and the photoresist destroyed. It is thought that the cause of the observed behaviour was the result of the Zn pellets being contaminated or oxidised during storage. With only a short time frame to complete the study and an order to new material likely to take several weeks the pursuit of Zn as a p-type dopant layer was deemed no longer an efficient use of time.

With the appropriate dopant no longer a viable choice and lack of adhesion a prominent issue it was decided that wetting layer would be required. Chromium, as used by Marburg was not available at the time of the study and so titanium was substituted, since it is widely used as such [140].

Titanium was only available for use with the JLS loadlocked sputterer (operational details of which can be found in section 3.5.3.2 of chapter 3) and so its use made contact deposition a two stage process, starting in the sputterer and then being continued in the evaporator (since the gold target of the sputterer is more expensive than the gold pellets consumed by the evaporator). The advantage of this method is that the sputterer is quick to load and pump down, whereas it can be several minutes between native oxide etch and achieving a vacuum with the evaporator. Thus, the chance of a native oxide causing adhesion issues is reduced.
150 Å of Ti was deposited, followed by a 260 Å cap of gold in order to give the evaporated gold the best conditions for adhering to the sputtered metals. Further to this ∼2000 Å of Au was evaporated, lift-off was semi-successful with the p-metal adhering well but the Au covering the masked areas not being totally removed, as is shown in figure 6.30.

In order to attempt to remove the gold left in place and to test how well the sputtered metal had bonded to the surface the sample was again soaked in acetone, but this time placed in an ultrasonic bath for an hour. The ultrasonic bath did remove a little more of the gold from the masked areas, but in doing so caused the evaporated gold to partially peel away from the sputtered gold beneath, with bubbles (presumably formed during lift-off) having formed between the layers of Au where it has not been entirely removed. Also, on the right of figure 6.30 droplets of gold can be seen to have formed across the p-contact during evaporation. The achievement of strong adhesion shows that the sputtered metal has produced a mechanically strong bond, albeit for an impractically thin layer of gold but that following with evaporation of a subsequent layer of gold does not work well since two independent regions of gold appear to be formed rather than one individual lattice.

Since the sputterer achieved good results, the cost of consuming the Au target was deemed reasonable if it led to high quality contacts. All further p-type contacts were produced using the sputterer only, with a final contact composition of 15 Å of Ti followed by 2000 Å of Au. An example of the contact quality produced using this approach can be seen in the left hand image of figure 6.29.

The n-type contacts used at Marburg were to be reproduced initially since they had been found to be superior to the p-type contacts and lacked major issues such as the observed burning. The Cr wetting layer was excluded partially because this material was unavailable but also since the optimised processing approach
bonds to BGaP as opposed to Si. The use of AuGe to form Ohmic contacts has been long established [107] and when applied to III-Vs does not require a wetting layer. It is likely that Cr was added in the Marburg process since the deposition was carried out onto Si, where the chemistry of the bonds is somewhat different.

The contacts were deposited using pellets of a eutectic alloy of AuGe in tungsten boats followed by a capping layer of pure gold to achieve a good ohmic contact that would be easy to bond gold wires to. AuGe was deposited to a thickness of 1040 Å and followed by a further 1000 Å of pure gold as measured by the thickness sensor, making the n and p contacts of comparable thickness. Lift-off was successful, with all metals deposited on the mask being removed by soaking in acetone and the n-contacts remaining attached. On inspection under an optical microscope (figure 6.31) it was found, however that although the n-type contacts had adhered to the BGaP surface they had formed a rough surface with many patches of defects throughout. Within the locality of the defects the Au capping appeared not to have adhered properly, leaving holes in the contacts.

The way in which Ohmic contacts are thought to form when using AuGe is that a eutectic fluid of AuGe at the metal/III-V interface dissolves the III-V causing Au-Ga to form, leaving the III-V slightly positively charged and the metal slightly negatively charged. This in turn causes the diffusion of Ge into the III-V where is acts as a dopant creating an \( n^+ \) region near the III-V surface, counteracting the Schottky barrier and allowing Ohmic contacts to form[107]. This does however assume that the P atoms stay in place, when in reality they could form either GeP or AuP, both in place or after having diffused some distance. The formation of either of these semiconductors within the contacts will act as defects in the ohmic contacts and may be the cause of the poor quality of the metalisation observed here.

![Figure 6.31: The defects produced when AuGe contacts were deposited by evaporation onto BGaP.](image)

The issue was rectified by adding a thin layer of nickel beneath the AuGe and between the AuGe and Au.
The addition of Ni has a two fold effect on the contacts, firstly it acts as a wetting agent, preventing surface tension from causing the metals to create three dimensional features rather than a plane of AuGe. Secondly it has been proposed that the addition of Ni causes the P (or As in the case of GaAs) to preferentially bond with the Ni, which diffuses into the semiconductor preventing defect formation within the contacts [119]. Adding Ni wire to a basket holder at a third electrode of the evaporator gave a final n-contact deposition structure of 40 Å of Ni, followed by 1000 Å of AuGe followed by another 40 Å of Ni and finally 1000 Å of Au. Lift-off was once more successful yielding contacts which under visual inspection (see figure 6.32) were defect free and shiny, suggesting a successful deposition.

![Figure 6.32: The un-annealed deposition of n-type contacts onto BGaP consisting of Au/AuGe/Ni.](image)

Having produced mechanically strong with good surface quality under optical microscope the final step was to determine the electrical properties of the contacts and optimise them by annealing. Two samples were metalised, one un-etched with p-type contacts and one etched down to the BGaP with n-type contacts. The probes were placed between adjacent p or n contacts (contact separation ~1 mm) and connected to a Keithley 2400 SMU to produce CW IVs, which are plotted in figure 6.33 along side the IV of the probe station characterisation system alone. Subtracting the series resistance of the system from the gradient of the IVs yielded p and n-type total resistances of 2.3 Ω and 1.31 Ω respectively. Assuming uniform current spreading the total resistance will be given by the following relation:

\[ R_{total} = 2R_{contact} + R_{bulk} \]  

\[ R_{contact} = \frac{\rho_{contact}}{A} \]
\[ R_{\text{bulk}} = \frac{\rho_{\text{bulk}} L}{A} \]

The contact resistance \( R_{\text{contact}} \) is defined as the impedance of the junction between metal and semiconductor (factor of two signifies two contacts required to make the circuit) and the bulk resistance \( R_{\text{bulk}} \) is the resistance of the semiconductor material separating the contacts. The separate resistances then can then be defined in terms of the relevant resistivities, \( \rho_{\text{contact}} \) based on the specific semiconductor/metal junction and \( \rho_{\text{bulk}} \), an intrinsic material parameter. For the two relevant semiconductors, BGaAsP and BGaP it is assumed that the boron content has negligible impact on the resistivity, allowing these to be taken from literature for GaAsP and GaP with values of 0.06 \( \Omega \text{cm} \) (calculated by interpolation of GaAs and GaP values) [71] and 0.02 \( \Omega \text{cm} \) [45] respectively for a doping level of \( 10^{19} \text{ cm}^{-3} \).

For a contact separation of \( \sim 1 \text{ mm} \) the bulk resistances for BGaAsP and BGaP are approximately 0.6 \( \Omega \) and 0.2 \( \Omega \) respectively. The contact resistances and resistivities, given by rearranging equation 6.9 for BGaAsP and BGaP will be equal to 0.85 \( \Omega \) and 0.56 \( \Omega \) and 850 \( \mu \Omega \text{cm}^2 \) and 560 \( \mu \Omega \text{cm}^2 \) respectively. The un-annealed contact resistivities, although large compared to commercially available lasers (<10 \( \mu \Omega \text{cm}^2 \) [7]) seem a reasonable starting point considering the successful formation of Ohmic contacts is reliant on the diffusion of dopants.

Figure 6.33: Contact resistances before and after rapid thermal annealing (RTA).

Since the Marburg processing procedure did not include a contact anneal (alloying) step it was required to develop an anneal from scratch. A RTA recipe used for AuGe/Ni as well as Au/Ti contacts was suggested by regular users. The suggested recipe was to ramp to 380 °C over 10 s, remain stable at 380 °C for 30 s.
and, since the RTA furnace did not feature active cooling to then let the temperature drop back to ambient naturally. The full details of the operation of the RTA furnace are presented in section 3.5.3.3 of chapter 3.

Contacts were repeatably produced with a smooth, shiny surface and no immediate signs of damage, contamination or defects. These samples were then electrically characterised once more in accordance with the method detailed above and the data added to figure 6.33 in the form of the open points. The addition of the annealing step was found to reduce the total resistance from 2.3 \( \Omega \) and 1.31 \( \Omega \) to 0.91 \( \Omega \) and 0.65 \( \Omega \) for the p and n-type contacts respectively. This translates to contact resistances and resistivities of 0.31 \( \Omega \) and 0.45 \( \Omega \) and 310 \( \mu \Omega \text{cm}^2 \) and 450 \( \mu \Omega \text{cm}^2 \) respectively.

The improvement is attributed to the thermally assisted diffusion of dopants and lattice defects. The reason for the p-type contacts being slightly more resistive than the n-type is thought to be related to the relative roughness of the BGaAsP capping in comparison to the etched BGaP, this could be investigated further by carrying out a short surface etch of a few nm to polish the surface across the entire, unmasked sample before definition of the mesa structure. The contact resistivities after annealing are significantly improved, more so for the p-type contacts than the n-type, which may be linked to the rough BGaAsP surface causing defects in the gold lattice which are subsequently diffused when thermally activated by the RTA process.

The final version of the wafer sample prepared using the optimised processing methodology is shown below in figure 6.34, where some roughness to the gold surface can be seen in patches (left) but is generally found to be smooth and shiny as can be seen by the zoomed view (right). To fit the time restraints of the study the processing, although having scope for further improvement was deemed a significant improvement on the previous process thus far and that some final characterisation should be carried out in order to allow direct device IV and LI comparisons to devices produced using the Marburg processing methodology.
6.2.2.4 Cleaving  To allow characterisation of the broad area laser structures to take place it was required to first cleave the wafer pieces into bars and in doing so define cavities. Cleaving silicon however, is somewhat challenging due to its brittle nature. Cleaving wafer samples into pieces of 1 to 2 cm\(^2\) for processing is fairly reliable, although the edges are generally not vertical or straight. Whilst fine for the edges of a sample to be processed (the edges are always of no use as a result of the edge bead, deposited metal wrapping the sides etc) it is far from suitable for defining cavity facets.

In order to achieve straight, vertical cleaved edges the Si substrate was first thinned to increase the probability of the break in the Si following the line of cleavage of the III-V region. The initial thickness of the samples was measured to be 780±10 μm consistently using a thickness meter. Deciding how much material to remove is a question of balancing the ease of cleaving the wafer piece into bars against making the bars themselves too fragile to handle. Previous samples received from Marburg that had been cleaved into bars were measured and it was discovered that the minimum thickness was ∼250 μm. These samples remained strong enough to be handled with tweezers and mounted in systems such as the closed cycle cryostat as required for characterisation studies. Thus, this thickness was chosen as a starting point to give the best chance of achieving straight cleavage lines and reasonable quality facets. If the samples remained challenging to consistently cleave at this stage further thinning could be employed as necessary at the risk of making the sample too fragile.

The procedure for thinning was to first apply a protective layer of photoresist to the metalised top side of the wafer sample, bake this to harden it and follow up by flooding on further photoresist and placing
the inverted sample on a lapping wheel sample holder. Baking the holder on a hot plate then hardened the photoresist allowing it to act as a resin to hold the sample in place and protect the top surface whilst thinning takes place. The sample holder consists of a circular inner section onto which the sample is mounted with a threaded collar. Rotated the collar sets an offset between the ring and inner section. This offset determines the thickness to which the sample can be thinned before the outer ring prevents further removal of Si. Fitting the sample to the holder is depicted schematically in figure 6.35.

Figure 6.35: The sample holder used to thin the Si substrate in tandem with lapping film. The threaded collar is adjusted to define the thickness of material to remove ($\Delta Z$).

Initially the sample holder was manually moved in a figure of eight pattern on a piece of 20 µm aluminium oxide lapping film, which was in turn placed on a glass plate. After approximately an hour lapping the sample the Si surface was visibly polished but little material had been removed. A second, more accurate thickness meter showed that approximately 10±1 µm of material had been removed, indicating that thinning to the target thickness would take approximately 53 hours. The more aggressive approach of using a motorised lapping wheel was therefore applied with a rotation speed of 400 RPM and lubricated with mineral oil.

A hour of lapping, moving the sample holder in a figure of eight pattern across the motorised lapping wheel, onto which the lapping film was mounted, thinned the sample by a further 25 µm. In addition to taking an impractical length of time the lapping film was being worn out at a rate of approximately 3 sheets an hour. As a final attempt to thin the samples 400 grit wet and dry sand paper was placed on the lapping wheel and the process repeated. The rate of thinning increased dramatically to $\sim 200$ µm per hour, including the time taken to replace sheets of sand paper, which were being worn in approximately 15 minutes. The sanding left the under side of the sample with deep scratches, which were a source of concern since the wafer may cleave along these rather than the surface scratches intentionally applied after thinning. Thus, once a
thickness of 260 μm was reached the sand paper was replaced with the 20 μm lapping film once more and a final thinning procedure carried out using the lapping wheel until all visible scratches were removed, leaving a final thickness of 250 μm.

Cleaving was carried out in the standard manner of scratching the device surface using a diamond scribe and cleaving along this line by attaching the sample to a backing film of self adhesive plastic (to prevent cleaved bars being dropped and lost) and using a razor blade as a pressure point. It was found that with this approach bars of length 1 mm could be produced with a success rate of ~75%. Given the fairly cumbersome nature of cleaving this was deemed reasonably successful and not pursued further due to the limited scope for further investigation as the result of time constraints. Further investigation could be carried out to determine if additional thinning would increase the yield of bars without compromising the strength of the devices significantly.
6.2.2.5 Optimised process flow  The final optimised process flow is presented here to give a clear picture of the overall methodology.

Figure 6.36: The optimised process flow developed in this study. Details are given in the relevant sections of 6.2.2, with the operation of equipment discussed in chapter 3.
6.2.3 Characterisation

With the remaining project time several of the devices processed into bars were characterised using the same probe station set up as that which was used with the devices sent from Marburg. By doing so a direct comparison between devices of the same structure (#25486 / #16195) differing only in growth run and processing methodology could be made. CW IVs were produced using the same probe station arrangement and Keithley 2400 SMU and are plotted along with the device of structure #25486 shown in appendix D, in figure 6.37.

The devices retain the same gradual turn on behaviour as those produced at Marburg but do so slightly more rapidly. This difference is fairly minimal and may well simply be the effect of the difference in growth runs between the two batches. However, of the devices tested from the processed batch the dispersion of the IVs was far less, with all devices giving comparable results. The Marburg processed devices on the other hand exhibited a large spread in performance, with many simply burning under electrical injection. This may be a reflection of an improved degree of consistency across devices using the optimised processing methodology.

The IVs shown here are approximately as good as those of the best of the Marburg processed devices but significantly superior to the worst, thus a larger sample of devices of sufficient quality for characterisation may be produced in this manner increasing the chance of finding a lasing device. One significant improvement was the ability to push the devices to the greater currents required to achieve lasing in the past without burning the contacts.

![Figure 6.37: The CW IV characteristics produced for four processed devices along side the IV of the device investigated in the ideality factor study with the same structure.](image-url)
With the optimised processing methodology the pulsed voltage source could be allowed to reach its limit of 100V, supplying currents in excess of 1 A without damage to the contacts. Having achieved this using pulsed injection the CW IVs were also allowed to reach greater currents, which again resulted in no damage to the bulk of the device or the contacts. In pushing the devices harder it was found that the devices produced using the optimised processing methodology did in fact turn on at $\sim 6.8$ V, as is indicated by the change from a linear to logarithmic trend in the log(I) vs V plot shown in figure 6.38.

Being able to apply a large enough bias to reach a more definitive turn on opens further avenues for more in depth study of the electrical characteristics of the devices. The increased gradient of the IV plot, indicating lower overall resistance could be related to an earlier onset of turn on and thus to growth differences between wafers; but would also be influenced by the contact resistance and bulk resistance of the device. The contact resistance may well be improved by the optimised processing procedure since the addition of an anneal step reduced this.

Furthermore the placing of the n-type contact on the highly doped BGaP should produce a better Ohmic contact than deposition on Si, leading to a reduced series resistance. Thus, it is reasonable to conclude that the electrical properties of the devices have generally been improved to a considerable degree through the optimisation of the processing methodology, but that the main limiting factors of the GaNAsP based lasers are more fundamental and will require attention at the growth stage.

Figure 6.38: Log scale IV of data plotted in figure 6.37. The removal of the contact burning issue allowed larger injection currents to be applied and thus achieve turn on at $\sim 6.8$ V.

Pulsed injection LIIs were produced using the probe station system discussed in sections 3.1 and 3.2.1 of chapter 3, with a pulse with of 500 ns and frequency of 10 kHz to limit Ohmic heating. The data for the
same four devices as presented in figures 6.37 and 6.38 are plotted in figure 6.39 and once more the data for a device of structure #25486 is included for reference purposes. The differences in external efficiency and output power at a given current are fairly minimal, again indicating good consistency across the batch of processed devices.

Comparing the LIs of the newly processed devices to the device of structure #25486 it is clear that a significant improvement in light output has been achieved through the optimisation of the processing. The improvement in output power could be related to facet quality, since this has been an area studied in the past, although it was found to have little impact on lasing performance [66]. The lateral etching of the unoptimised process was a point of concern, since it had been claimed that this was taking place to a significant degree for the QW(s). This was consistent with observations of capping layer etching and the large difference in etch rate of high As fraction compared to high P fraction materials in aqua regia.

With the optimised processing method it has been shown that the etch rate for BGaAsP and BGaP is fairly comparable (figures 6.23 and 6.24) and there is no longer visible evidence of lateral etching of the BGaAsP capping. From both of these results it can be inferred that the lateral etching of the QW would also have been reduced to negligible levels if it was indeed taking place to any significant degree previously. Therefore with a larger active area being pumped it would make sense that the optical output power would increase as is observed here.

Lasing was not observed even when raising the injection current to over an amp or cooling the devices to cryogenic temperatures. The major limiting factors such as defect related recombination and carrier leakage detailed in chapter 5 have not yet been successfully addressed. Thus the comparatively minor, although not insignificant improvements to the processing procedure are not expected to improve performance enough to achieve lasing alone.
6.2.4 Processing conclusions

The development and application of an optimised processing procedure for fabricating bars of GaNAsP based laser and LED structures from as-grown wafer material has been presented. Lithography, etchant choice and application, metalisation, contact annealing/alloying, thinning and subsequent characterisation of devices has been addressed and in each case led to an improvement in results over the previous methodology. The use of AZ5214E photo resist over the thicker AZ4562 thick-film resist led to a lesser edge bead and as a result allowed accurate reproduction of mask features down to sub micron resolution.

In choosing this less aggressive etchant the alternative photoresist, AZ5214E became a viable choice in terms of both achievable mask pattern reproduction and protection level for the sample from the etchant. The rough contacts of un-annealed Cr/Au and Cr/AuGe (p and n-type respectively) that were frequently burned during electrical injection were investigated.
The wetting layer was replaced with Ti purely out of availability since both are accepted wetting materials, however since this metal was only available in the form of a sputter target use of the slower and more expensive sputterer was required for the p-type contact. Further, the adhesion of evaporated gold to sputtered gold was found to be poor and so it was necessary to form the full contact using the sputterer only. This did however lead to contacts which were both mechanically strong and Ohmic.

The n-type contacts when reproduced according to the previous approach were found to adhere to the semiconductor surface but contained defects. The addition of Ni as both a wetting layer and an alloying component for the remaining P, after Ga was dissolved into the AuGe liquid led to smooth, shiny and strong contacts of structure Ni/AuGe/Ni/Au. The addition of an annealing step for these contacts was found to reduce contact resistivity from 850 $\mu\Omega \text{cm}^2$ and 560 $\mu\Omega \text{cm}^2$ to 310 $\mu\Omega \text{cm}^2$ and 450 $\mu\Omega \text{cm}^2$, a reduction of 64\% and 20\% for the p and n-type contacts respectively.

The devices were thinned using a mechanical lapping wheel and 400 grain sand paper followed by 20 $\mu$m aluminium oxide lapping film to remove scratches and polish the surface. The characterisation of the devices indicated evidence of reduced series resistance, the elimination of the contact burning problem, increased optical power and potentially improved facet quality.

It is interesting to note that the lateral etching is comparable in devices of structure #16195 (see figure 6.29) to those of structure #25486 shown in the SEM images above, yet still exhibits the best performance of any measured to date. From this fact it is inferred that the major limiting factor of the newer devices is not the quality of processing but that improvements in this area remain essential in preventing phenomena such as contact burning, which inhibit experimental characterisation.

Were wafer material of quality comparable to that of #16195 available it is likely that reasonable improvements in lasing efficiency could be achieved through the processing optimisations discussed here. Thus, although these improvements alone will not lead to achieving room temperature electrical injection lasing of GaNAsP based lasers. They will allow the necessary experimentation required to achieve this result possible, and are required to achieve a commercially viable product once the fundamental issues have been addressed.

6.3 Chapter conclusions

In this chapter three approaches to improving the GaNAsP/Si laser devices have been presented. By maximising the optical confinement factor via passive waveguide simulation, producing an accurate model of the electrical properties of the devices based on a modified version of the ideality factor equations and finally by the optimisation of the procedure for fabricating laser devices from as-grown wafer material.

The waveguide simulation study concluded that an improvement in confinement factor from 0.35 \% to 0.6 \% and 1.33 \% to 1.73 \% for single and multiple QW structures respectively could be achieved through minor adjustments to the cladding and barrier thicknesses. By increasing the confinement factor the modal gain is increased, leading to a reduced threshold current density and lower fractional contribution of highly carrier density dependent non-radiative recombination mechanisms such as carrier leakage, which has been demonstrated to be a prominent efficiency limiting process in these devices.
Furthermore, the prevention of the optical mode from coupling into the p-doped cladding reduces the rate of free carrier absorption, which has also been indicated to limit the GaNAsP-based lasers. This approach to improving the confinement factor has the benefit of requiring no major structural changes, only minor adjustments to layer thicknesses and as such the impact on other parameters is minimal. However, the index contrast has a far larger impact on confinement than layer thicknesses and as such offers an additional approach to improving modal gain.

The refractive index though is inversely proportional to the band gap meaning that increasing one will decrease the other, causing a balance between maximising optical and carrier confinement. In addition, material changes will also alter strain compensation and the thermal expansion coefficients potentially causing increased strain fields, defect density and potentially even cracking. It was shown that small changes in barrier composition cause only minimal improvements in index contrast, suggesting that this may not be easily increased for this material system.

The processing optimisation study successfully addressed the key limitations of the previous processing methodology, namely contact burning, lateral etching, rough and inconsistent contacts and etching inaccuracies. These were achieved by firstly optimising the choice of photoresist, exposure and developing parameters to achieve sub-micron lithography resolution. The etchant was then modified with the addition of acetic acid to reduce the etch rate, improve the resultant surface quality and reduce lateral etching to negligible levels.

The metalisation was also investigated leading to the use of sputtered Ti/Au p-type contacts and evaporated Ni/AuGe/Ni/Au n-type contacts, which were found to be Ohmic and defect free. The application of an annealing processes to the contacts led to a 64% and 20% reduction in contact resistivity over the un-annealed contacts. Lastly devices were cleaved into bars after thinning the wafer from 750 μm to 250 μm by lapping, producing broad area lasers of cavity length 1 mm.

The characterisation of these lasers indicated that although lasing was not achieved, the electrical properties had been somewhat improved, potentially through a reduction in contact resistance, and that with contact burning eliminated the devices could have much larger injection currents applied. Electroluminescence provided evidence of increased optical output power and efficiency, which is thought to be related to an increase in pumped QW area through decreased lateral etching and a potential improvement in facet quality.

The net result of the processing study is a general improvement in device quality, which although not addressing the fundamental physical limitation of strong defect-related recombination and carrier leakage allows experimental characterisation to take place more easily, increasing the scope for carrying out studies which will determine how best to approach these issues.

The combined results of this chapter represent two alternative approaches to improving the GaNAsP-based lasers in a situation in which standard experimental methods are not applicable due to a lack of functional devices. It is hoped that these studies will allow the fabrication of devices of sufficient quality to for experimental characterisation to take place once more and to supplement these studies with additional scope for analysis.
Chapter 7

7 Characterisation of GaSb quantum well lasers for monolithic integration on silicon

7.0.1 The background of III-Sb growth on silicon

The large lattice and thermal expansion coefficient disparity between III-Vs and Si has been overcome in two different ways so far in this thesis by developing a novel lattice matched material (the GaNASP based lasers discussed in chapters 5 and 6); or by fusing a separately grown InP based active region to the Si substrate and inhibiting the propagation of threading dislocations with the inclusion of a defect blocking layer (the fused wafer devices of chapter 4). An alternative is to grow a buffer layer of a material that can accommodate a greater strain than usual through the preferential generation of an alternative type of defect, which is immobile in the direction of slip of the crystal. In doing so the defects will tend to form close to the III-V/Si interface, releasing the strain and allowing further epitaxially deposited material to be formed free of excessive strain and the associated defects. For the industrial standard laser materials based on GaAs and InP this is not the case, with the preferential generation of threading or screw dislocations which propagate in the growth direction, leading to the large defect densities and poor performance at early attempts to integrate such materials on Si [41]. III-Sb shows an unusually large capacity for the strain that results from large lattice constant differences, for example as early as 1986 GaSb photodetectors were successfully realised on Si substrates through the use of an AlSb buffer layer, a system for which the lattice constant mismatch is as much as \( \sim 11.5\% \) [15]. The GaAs/Si mismatch is a comparatively small \( \sim 4\% \) at 300 K, indicating huge potential for III-Sb materials to be integration on Si. Work on AlSb as a buffer layer has been continued by the group of Huffaker, where the ability of AlSb to accommodate such large lattice mismatches is demonstrated to result from the nucleation layer forming AlSb quantum dots by Stranski–Krastanow growth, which after further deposition coalesce to form a undulating structure, increasing the surface area and therefore allowing greater strain to be accommodated [60].

The group of Tournie at Montpellier University have also approached the issue of photonic integration on Si through the use of III-Sbs, in this case using a GaSb buffer layer, which accommodates the strain in a slightly different manner. GaSb is able to accommodate the large quantities of strain originating from the lattice constant and thermal expansion coefficient mismatch due to its relative softness and the fact that defects tend not to propagate. Thus a high defect-density layer near the interface will form, relaxing the strain and allowing high quality, unstrained and defect free material to be epitaxially deposited thereafter [99]. The displacement of atoms as a defect propagates through a lattice is defined by the Burgers vectors, which for FCC materials such as GaSb are the variations of [110], since these minimise the energy of the defect state [65]. The result is that for growth in the [100] direction (for CMOS compatibility) defects will propagate at 60° to the surface. As a result the defects will intercept each other and, again from the minimisation
of energy will form what is known as a Lomer-Cottrell lock; whereby the resultant Burgers vectors prevent dislocation propagation in the [100] direction, allowing defect free material to be formed by further epitaxy [95].

Strained $Ga_{0.8}In_{0.2}Sb$ QWs with $Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$ barriers were grown on GaSb substrates as proof of concept initially and found to lase at 1.55 $\mu m$ up to 45 $^\circ C$ with a threshold current density of 330 $Acm^{-2}$ per QW followed by the same structure on a Si substrate, lasing at room temperature with a threshold current density of 5 $kAcm^{-2}$ [90] (1.67 $kAcm^{-2}$ per QW). More recently an optimisation of the GaSb substrate lasers using GaInSb/AlInSb composite quantum wells (see figure 7.3 for details) has led to CW electrical injection lasing at room temperature with a threshold current of 85 mA and a 30 mW/facet emission power for 10 $\mu m$ by 1 mm dimension lasers [89]. The impressive performance of these devices for a relatively new approach to III-V integration on Si suggests strong potential for the GaSb based laser approach in realising a commercially viable silicon compatible laser.

### 7.0.2 Montpellier GaSb laser structure details and development

In this chapter the characterisation and physical limitations of three different structures of the strained $Ga_{0.8}In_{0.2}Sb/Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$ QW lasers grown on GaSb substrates are presented and discussed. The three different structures vary only in QW composition, with the layers remaining as three $Ga_{0.8}In_{0.2}Sb$ QWs, $Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$ barriers, $Al_{0.9}Ga_{0.1}As_{0.07}Sb_{0.93}$ cladding and GaSb substrate, as depicted in figure 7.3 left with the full layer details tabulated in 7.2 right.

<table>
<thead>
<tr>
<th>Material</th>
<th>Thickness (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-GaSb(100)</td>
<td>300</td>
</tr>
<tr>
<td>Graded p-AlGaAsSb</td>
<td>100</td>
</tr>
<tr>
<td>p-$Al_{0.9}Ga_{0.1}As_{0.07}Sb_{0.93}$</td>
<td>1000</td>
</tr>
<tr>
<td>$Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$</td>
<td>200</td>
</tr>
<tr>
<td>$Ga_{0.4}In_{0.6}Sb$</td>
<td>3.6 (A), 4.8(B), 6(C)</td>
</tr>
<tr>
<td>$Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$</td>
<td>20</td>
</tr>
<tr>
<td>$Ga_{0.4}In_{0.6}Sb$</td>
<td>3.6 (A), 4.8(B), 6(C)</td>
</tr>
<tr>
<td>$Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$</td>
<td>20</td>
</tr>
<tr>
<td>$Ga_{0.4}In_{0.6}Sb$</td>
<td>3.6 (A), 4.8(B), 6(C)</td>
</tr>
<tr>
<td>$Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$</td>
<td>200</td>
</tr>
<tr>
<td>n-$Al_{0.9}Ga_{0.1}As_{0.07}Sb_{0.93}$</td>
<td>1000</td>
</tr>
<tr>
<td>Graded n-AlGaAsSb</td>
<td></td>
</tr>
<tr>
<td>n-GaSb(100)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 7.1: Left: The device structure for all GaSb based structures investigated in this study as well as [90] (image taken from this paper) and [89]. Right: Tabulated device composition details.

The design of the structure in figure 7.1 starts with the requirement that the materials used lattice match to GaSb, since this is the material which allows monolithic growth on silicon to take place as discussed in
section 7.0.1. To form an SCH type structure that will give both strong optical and carrier confinement three materials are required, one of large band gap for the cladding, one intermediate bandgap for barriers and one of smallest band gap (must be direct) that can be used to form a QW. The Quarternary AlGaAsSb is suitable for both barriers and cladding since by varying the Al content a large change in band gap can be achieved for minimal lattice constant increase from that of GaSb, making lattice matching simple. This is shown in figure 7.2, where it can be seen that the additional of a small fraction of As (causing a movement towards GaAs) counteracts the increase in lattice constant with added Al. This material, for Al fractions of more than $\sim 30\%$ causes the quarternary to become indirect, thus making the material unsuitable for photon generation and only for barriers or cladding.

Figure 7.2: Bandgap vs lattice constant, showing how the binaries GaSb, AlSb, InSb and GaAs can be combined for form an SCH structure lattice matched to GaSb. Image taken from [142].

The QW could be formed from pure GaSb with sufficiently thin QWs causing enough of an increase in confined state energy to allow emission at the $\sim 0.8$ eV energy corresponding to a 1550 nm lasing line. However if this is done the barriers and QWs will lattice match, causing no strain to be generated and the benefits of such strain foregone. Increasing the uniaxial compressive strain incorporated into the QW causes the x-y plane (if z is taken to be the growth direction) leads to a reduction in hole effective mass in this plane [11]. This reduction, in accordance with equation 2.46 leads to a lower density of states and therefore carrier density required to reach population inversion [85]. The compressive strain of the QW can be increased simply by choosing a material of larger lattice constant than the barriers according to the relation $\epsilon_|| = \frac{(a_s-a_{ep})}{a_{ep}}$ where $\epsilon_||$is the uniaxial strain in the x-y plane, $a_s$ is the substrate (in this case barrier) lattice constant and $a_{ep}$ is the epilayer (QW) lattice constant.

With reference to figure 7.2 once more it can be seen that the addition of indium to form GaInSb is the only suitable choice for increasing the lattice constant whilst maintaining a suitably large conduction band offset relative to the AlGaAsSb barriers. The calculated bandgap as a function of indium content is shown.
in figure 7.3 for both bulk (black) as well as the ee1 to HH1 transitions of the strained QWs for the three structures investigated in this thesis (see figure 7.4 for details). The ee1 - HH1 transition has an increased energy over the bulk bandgap from quantum confinement (see section 2.12.1) with the energy increase from the bulk CB edge to ee1 state (and similarly for the valence band) given approximately be equation 2.45 or more precisely by numerically solving the transcendental equation for a well of finite depth (as was done by in-house simulation here with the assistance of Dr. S. Jin). The transition between between the ee1 and HH1 levels of the QWs is achieved for an indium fraction of \( \sim20\% \), which in addition yields a compressive strain of 1.24%, 1.3% and 1.35 for structures A, B and C respectively when the composite well structure shown in figure 7.4 is used.

Figure 7.3: The band gap of GaInSb as a function of indium fraction for bulk as well as the varying strained QWs detailed in figure 7.4.

QW composition A (2072) is the 3.6 nm well that was grown within the structure of 7.2 and investigated in [90], representing a basic triple quantum well laser as a base line. QW composition B (2073) is a composite well of overall (nominal) width 5.25 nm consisting of two 2.4 nm \( Ga_{0.8}In_{0.2}Sb \) layers separated by a single monolayer (0.45 nm (nominally)) of \( Al_{0.68}In_{0.32}Sb \). QW composition C (2079) is also a composite well but in this case has an overall width of 6.9 nm formed from three 2nm \( Ga_{0.8}In_{0.2}Sb \) layers separated by two single monolayers (0.45 nm (nominally)) of \( Al_{0.68}In_{0.32}Sb \).

Increasing the QW thickness from A to B to C increases the gain volume, which would be expected to reduce the threshold carrier density by lowering the required band filling and increasing the photon generation rate for a given injection current. In addition the increased optical confinement factor would also be expected to contribute to the reduction in threshold carrier density by increasing the modal gain and
therefore stimulated emission rate. A reduction in carrier leakage is expected as a result of lower band filling and therefore higher effective confining potential. Defect related recombination would likely remain relatively unchanged and Auger recombination also be reduced due to lower band filling (carrier density) and reduced internal heating of the device as lower threshold injection current is required.

The benefits of the composite quantum well structure are twofold. Firstly additional compressive strain can be included into the well by the lattice mismatch between the $Al_{0.68}In_{0.32}Sb$ monolayers and GaInSb leading to a slightly larger bandgap (going some way to counteract the reduction caused by the wider well) as well as a further reduction in density of states. Secondly since the layers are so thin the tunneling rate will be high (allowing carriers to spread easily throughout the well) and the electron wavefunctions of the sections will overlap creating coupled quantum wells, similar to a superlattice. The coupling of the wells has a somewhat analogous effect to the difference between the energy levels of an isolated atom to a crystal lattice, all be it to a lesser extent since there are only three coupled wells. The lowest excited states (ee1 and HH1) are therefore split into three energy levels that cause an effective increase in bandgap of a few meV.

![Figure 7.4: The three QW structures investigated here. A: 3.6 nm $Ga_{0.8}In_{0.2}Sb$ QW (identical to that studied in [90]). B: 4.8 nm $Ga_{0.8}In_{0.2}Sb$ QW split by a monolayer (0.45 nm) $Al_{0.68}In_{0.32}Sb$ barrier. C: 6 nm $Ga_{0.8}In_{0.2}Sb$ QW split by two monolayer (0.45 nm) $Al_{0.68}In_{0.32}Sb$ barriers.](image)

This chapter discusses the temperature and hydrostatic pressure dependent electroluminescence studies of the GaSb based strained QW lasers and the key physical processes limiting device efficiency are identified and quantified and the relative performance of the three structures contrasted.
7.1 Temperature dependent characterisation

7.1.1 Ambient temperature electrical properties

The CW injection electrical performance was initially checked to see if the devices had any obvious features that may be of interest. Figure 7.4 shows the IVs for A, B and C with turn-ons of starting and ending between 0.8 and 1.3 V and static resistances of 35, 31 and 30 Ohms respectively. The differences between the three structures are negligible, as is to be expected since they vary only in QW composition and could reasonably be expected from three devices of even a single structure from device-to-device variation. Slight differences in connection quality between the probes and devices could even cause minor differences and so the IVs for each are treated as equal. Since these devices are designed to operate at 1550 nm, corresponding to 802 meV the turn on starting at ~0.8 V is, as would be predicted theoretically, consistent with the “ideal” commercial InGaAs laser discussed in appendix D. The difference here is that these devices comparatively smoothly turned on over ~0.5 V compared to ~0.1 V in the case of the ideal device. The increased voltage required to complete turn on is suggestive of the carriers being more dispersed in energy so that whilst at 0.8 V the highest energy carriers are beginning to flow across the junction in reasonable numbers it is not until ~1.3 V that close to 100% have sufficient energy to do so.

The static resistances (characterisation system impedance accounted for) are also somewhat large, compared to ~8 Ohms in the case of the ideal device. It should be noted that the static resistance has no effect on ideality factor (A factor representing how close to ideal a diode junction is, based on Shockley’s ideal diode equation [29]) or turn on behaviour and as such is not related to the properties of the junction. Therefore the conclusion drawn here would be that factors such as inherent material impedance and contact metal/junction quality would be the limiting factors, it should be noted that these devices were simply contacted top and bottom, thus requiring the current to pass through the full thickness of the substrate, increasing the overall resistance. Thus from an electrical perspective only, the devices all act as reasonably ideal diodes following the theoretically predicted behaviour pattern with only fairly minimal deviations.
7.1.2 Temperature dependent electroluminescence

Pulsed electrical injection LI characteristics were produced as a function of temperature from 40 - 300 K in steps of 20 K using the closed cycle cryostat system and presented in figure 7.6 for structures A, B and C (alternate temperatures omitted for clarity). Lasing was found to occur up to 300 K (the limit of the cryostat system) in all three structures with $J_{th}$ values of 1092 $A cm^{-2}$, 555 $A cm^{-2}$ and 471 $A cm^{-2}$ per device or 364 $A cm^{-2}$, 185 $A cm^{-2}$ and 157 $A cm^{-2}$ per QW for A, B and C respectively. The industry standard InGaAs/InP lasers exhibit current densities of around 100 $A cm^{-2}$ per quantum well [113] and so structure C, being only 50% greater is comparable and would not require too significant an improvement to be competitive.

This is of course on a GaSb substrate, on Si an increase in threshold current density would be expected since although GaSb is found to suppress threading dislocations levels of strain and defect density can be achieved that will overcome the Lomer-Cottrell lock. This is reflected in the published $J_{th}$ value of 5 $kA cm^{-2}$ (1.67 $kA cm^{-2}$ per QW) for structure C grown on Si [90], which is far greater than that of commercial lasers. Therefore any improvements in device efficiency that could be achieved on the GaSb substrate will not only carry over onto the Si substrate devices but also: a) prevent these loss mechanisms being exacerbated by factors such as the elevated threshold carrier density of the lasers on Si for example. And b) allow the losses directly associated with integration on Si to be more easily isolated from those inherent to the III-V material system and addressed appropriately.

Comparing the three structures it is clear that the change in QW structure from A to B to C is beneficial
with an initially large (almost halving) of $J_{th}$ (A to B) followed by a smaller, yet still significant decrease from B to C. This is indicative of a reduction in non-radiative recombination in accordance with equation 2.21. As was mentioned in 7.0.2 defect related recombination would be expected to be unchanged and so a reduction in Auger recombination or carrier leakage must be taking place, both of which would be consistent with the expected reduction in threshold carrier density from the increased gain volume (lower band filling), lower HH effective mass and increased modal gain (increased confinement factor).

Increasing the gain volume (A to B to C), although beneficial in terms of increasing the availability of states to be depopulated by stimulated emission concurrently and thus increases the photon density and output power for a given carrier density also requires a larger current to achieve such a carrier density. Thus, at some stage the reduction in threshold current density will begin to slow (and in fact may increase) with ever increasing gain volume (this is a consideration when choosing the optimal number of QWs for a given structure) and so this is likely part of the reason for a lesser improvement from B to C compared to from A to B.

Generally when considering a differing number of QWs this is somewhat taken into account by quoting a threshold current density per QW, however since the number of wells here is constant the threshold current density per nm of QW, although an unconventional metric may be of most relevance. The thickness of QW material is $(1 \times 3.6)\times3$ QWs = 10.8 nm, $(2 \times 2.4)\times3$ QWs = 14.4 nm and $(3 \times 2)\times3$ QWs = 18 nm giving threshold current densities per nm of 101 $Acm^{-2}nm^{-1}$, 39 $Acm^{-2}nm^{-1}$ and 26 $Acm^{-2}nm^{-1}$ for A, B and C respectively. When considered in this format we see that the changes in $J_{th}$ and $J_{th}nm^{-1}$ from A to B (B to C) are 49 % and 61 % (15 % and 33 %) and thus although appearing comparatively minimal from the LIs alone (15%) the improvement per nm of material added from B to C is actually still very significant (33%).

The slope efficiency, although not possible to absolutely calculate due to low collection efficiency from the facet using the bare fibre collection system from the cryostat (see section 3.3) can be investigated in terms of its trends and relative changes. In all three devices (with the possible exception of A at 300 K) $\eta_d$ appears fairly constant with temperature and a sharp change at threshold is observed as would be expected for idealised lasers. These results would imply reasonable pinning behaviour and a low contribution from optical losses.

The temperature dependence of both $J_{th}$ and $\eta_d$ is discussed in section 7.1.3 with reference to the characteristic temperature $T_0$ and $T_1$, however just by visual inspection a strong dependence on temperature is clear in all three devices with $J_{th}$ dropping to 18, 16 and 22 $Acm^{-2}$ for A, B and C respectively by 40 K.
Figure 7.6: Pulsed electrical injection LI characteristics collected from the facet over the temperature range 40 - 300 K using a closed cycle cryostat for structure A (2072).

An example of the integrated spontaneous emission collected from these devices as a function of temperature is shown in figure 7.9. For brevity an example for structure 2079 is shown since results were similar in each case. The integrated spontaneous emission, and therefore carrier density clearly pins from 300 K down to 160 K but below this temperature pinning is observed to some extent, at least initially, but the collected power increases with injection with an increasing rate for decreasing temperature. A failure to pin would lead to an increased contribution from highly carrier density dependent mechanisms, i.e. Auger recombination and carrier leakage, although both of these are also highly temperature dependent and should be comparatively minimal at low temperature.

However, if the carrier density is pinning (as it seems to be down to \( \sim 160 \) K) then because the spontaneous emission rate is proportional to \( n^2 \) (see equation 2.9) then a lower spontaneous emission power would be expected as the threshold current (and therefore threshold carrier density) falls. The opposite is observed here, completely counter to expectations, which would suggest a temperature dependent change in collection efficiency. Measurements were taken both decreasing and increasing temperature to account for processes such as vibrations or ice build up causing a change in alignment of the detection system and no such effect was observed. Changes in refractive index with temperature could alter the shape of the emission cone or affect the scattering of stimulated emission or amplified spontaneous emission reversibly and so something of this kind may be happening here. Whatever the effect it serves is leads to the spontaneous emission measurements somewhat unreliable and so they will not be used further in this investigation to avoid introducing an
additional unknown into investigations such as the Z analysis.

Figure 7.7: The integrated spontaneous emission as a function of temperature collected from the sides of the devices using a bare fibre due to completely metalised top and bottom.

7.1.3 Characteristic temperature analysis

Plotting the logarithm of the threshold current density and the slope efficiency against the temperature and taking the inverse gradient of each gives the characteristic temperatures $T_0$ and $T_1$ as shown in figure 7.8. The characteristic temperatures represent the temperature sensitivity of the threshold current density and slope efficiency and can be directly compared between other devices as well as to “characteristic” values representing the limits of recombination by each mechanism exclusively.

The inverse gradient of the black data (left axis) represents $T_0$, with average values across the temperature range of 63±6 K, 68±6 K and 92±6 K for A, B and C respectively. The characteristic temperatures in all cases are very low (compared to the ideal radiative limit of $T_0 = T$) in the range 100 K to 300 K, and below 100 K rather large. Figure 7.9 shows the characteristic temperatures calculated for three point groups (e.g. the point at 60 K is the inverse gradient from 40 K - 80 K) as a function of temperature along with the theoretical limits for radiative, defect and Auger dominated devices (the carrier leakage dominated limit is challenging to calculate) to clarify this discussion. These limiting values are taken to be (see section 2.10) $T_0 = T$ for a radiatively dominated device, $T_0 = \frac{24}{3}T$ for a defect dominated device and $T_0 = \frac{T}{3 + E_a/k_bT}$ for Auger recombination, where $E_a$ is the Auger activation energy and can be approximated to $E_a \approx 3k_bT$.

From this plot Auger recombination appears to be a limiting factor in all three devices, in agreement with the discussion above, pulling $T_0$ down to these low values and to a lesser extent in B and lesser still in C for the reasons detailed previously. As the temperature is lowered $T_0$ tends towards the defect and radiative
limits, as one would expect. Radiative recombination would be expected to be more dominant but the effect of defect related recombination is somewhat more difficult to discern from these results alone. A value close to the defect characteristic temperature may simply be the result of a combination of radiative and Auger recombination, making the defect contribution smaller than it seems. As these devices are lattice matched to their GaSb substrate defect densities are not expected to be large, however other sources of defects may be present that give some small (relative to the other recombination mechanisms) contribution.

The large $T_0$ value at very low temperatures can be an indication of an inhomogeneous transition energy within the active region (see figure 5.10 in chapter 5 section 5.2.1). If this varies spatially then at threshold some carriers may meet the Bernard-Duraffourg condition, whereas others have too great an energy to contribute to lasing. At low temperatures the states can become localised as carriers lack the thermal energy required to transition to lower energy neighboring states. As the temperature increases the localised carriers are increasingly able to escape into the lower energy states and contribute to gain, reducing the current that would have had to be injected were the band gap homogeneous and thus increasing $T_0$. This type of low temperature $T_0$ behaviour is also observed in quantum dots, since these tends to form with a range of sizes and behave similarly to an inhomogeneous active region. The cause in these devices could be the composite wells, which couple the wavefunctions of the sections and cause energy level splitting. It has been shown [61] that if the wells in a superlattice (an almost identical situation to the composite wells) are not identical localised states in the mini-bands form that act analogously to an inhomogeneous active region. This is a plausible cause of the unusually large $T_0$ value seen in these devices since the expected energy splitting is $\sim$5 - 6 meV [61], consistent with a thermal energy of 50 - 70 K, around the value at which $T_0$ is found to be too large.

The difference in $T_0$ between the three structures is consistent with the trend in threshold current density and highlights the improvement in performance through increasing the composite well width, most likely leading to a reduction in Auger recombination.
Figure 7.8: The log of the threshold current density and slope efficiency as well as $T_0$ and $T_1$ characteristic temperatures plotted as a function of temperature from 40 - 300 K for all three structures.

The temperature sensitivity of the slope efficiency, $T_1$ is somewhat different to $T_0$, varying strongly with temperature and from structure to structure. The $T_1$ values around ambient temperature are fairly low when compared to commercial devices where values of 1000K are attainable[59], (particularly for A) at 17±2 K, 84±9 K and 101±11 K for A, B and C respectively indicating a highly temperature sensitive slope efficiency. At low temperatures on the other hand the $T_1$ values are extremely large at 330±48 K, 751±60 K and 1472±74 K for A, B and C respectively, suggesting that at low temperature the slope efficiency is almost completely independent of temperature. The very different low and high temperature $T_1$ values suggest then that a highly temperature sensitive loss mechanism is activated at approximately 120 K that was comparatively negligible before but dominant thereafter.

With the slope efficiency defined according to equation 2.34 in chapter 2 there are two possibilities for the origin of the increasingly temperature sensitivity, optical losses ($\alpha_i$) such as IVBA and FCA, and the internal quantum efficiency. Optical losses are a possibility since an increased thermal spread of carrier energy will increase the cross section for re-absorption, as states between which a transition may take place whilst conserving momentum become occupied. The carrier excited by the absorption will then tend to decay via phonon emission, generating heat. This could also be tied into the increasingly large threshold current density and therefore carrier density required at larger temperatures, since this will also increase the proportion of states available for optical loss. If optical losses start to become prominent then the gain will be lowered, which will in turn require a larger carrier density to reach threshold gain, thus this interaction is a positive feedback loop that will become increasingly worse, explaining the sharp reduction in $T_1$ with temperature.
The other possibility is a temperature dependent reduction in the internal quantum efficiency, which is defined by equations 2.35 and 2.36 in chapter 2. This parameter is a function of the fraction of the radiative current compared to the total injection current attributed to all recombination mechanisms. Therefore at higher temperatures the more temperature dependent mechanisms, Auger recombination and carrier leakage will exhibit a fractional increase in their associated currents. In addition at low temperatures (below \(\sim 120\) K) where the \(T_1\) values are very large these mechanisms should be fairly negligible in comparison to the other competing recombination mechanisms. All data analysed thus far has pointed to these two mechanisms (Auger recombination in particular) being present to at least some degree, if not dominating recombination in these devices and the sharp drop in \(T_1\) being at least partly caused by a reduction in internal quantum efficiency as a result of these processes is therefore likely. This hypothesis is also supported by the increase in ambient temperature \(T_1\) value with increased composite quantum well thickness, particularly from structure A to B. The cubic and (approximately) exponential dependencies of the Auger and carrier leakage currents on the threshold carrier density lead to an expectation that the increased modal gain, reduced band filling and increased strain with increasing composite well width would strongly decrease the contribution of these mechanisms, fitting these observations.

Figure 7.9: The characteristic temperatures for all three structures as a function of temperature along with the theoretical limits for radiative, defect and Auger dominated devices.

7.1.4 Z analysis

Carrier leakage and particularly Auger recombination have thus far been indicated to be important recombination mechanisms in these devices, this hypothesis can be supported further by quantifying the approximate
carrier density dependence of the threshold current using the Z analysis. The details of how the Z parameter is calculated and its relation to the recombination mechanisms and associated currents is discussed in section 2.9.1 with reference to figure 2.15 in chapter 2. One problem with the use of the Z analysis in this instance is the somewhat unusual nature of the integrated spontaneous emission and the ambiguous causes there of. In order to obtain a more reliable estimate of the pure spontaneous emission as a function of injection current the very low injection, sub threshold facet emission where gain is low and stimulated emission is negligible is extrapolated to threshold. The extrapolation relies on the assumption that at low injection radiative recombination dominates and therefore \( I \approx I_{\text{rad}} \) and that the spontaneous emission rate is proportional to the radiative current, \( r_{sp} \propto I_{\text{rad}} = eVBn^2 \). Examples of the extrapolations from linear regions of the sub threshold LIs can be found in appendix E.

The Z parameter was then calculated as normal using the approximated spontaneous emission for each temperature and all three structures, and is plotted in figure 7.10 below. The Z values at ambient temperature are 2.91±0.08, 2.85±0.08 and 2.83±0.08 for structures A, B and C respectively, decreasing down to almost exactly 2±0.08 for all three at 40 K. The ambient temperature values are all close to 3, the characteristic value for an Auger dominated device which would again fit with the previous suggestions of a highly temperature dependent loss mechanism being present in all three structures. In addition, the gradient of the Z vs T plot is decreasing with temperature suggesting a possible asymptotic trend towards \( Z = 3 \), which would further support Auger recombination being a key non-radiative recombination mechanism. The reduction in \( Z \) with increasing composite well thickness would also be consistent with Auger recombination as discussed previously.

The low temperature value of 2±0.08 for all three structures indicates a proportional increase in radiative recombination as temperature is decreased, consistent with highly temperature dependent loss mechanisms, as was indicated above. The trend does not seem to tend towards 2 however, which could indicate that \( Z \) would decrease further if lower temperatures could be reached, implying the presence of defect recombination. That the three data sets coincide at low temperature would again be consistent with highly carrier density and temperature dependent loss mechanisms, since these will be reduced for increased composite well thickness, whilst defect related recombination would be expected to be consistent across all three structures. Carrier leakage has also been suggested as a possible key loss mechanism in these devices but is somewhat difficult to identify via Z analysis unless Z takes values outside those achievable by a combination of defect, radiative and Auger recombination. This is the case because the characteristic value for a carrier leakage dominated device is a function of the confining potential for the carriers within the QWs. This parameter varies with band filling, and active region composition, making it somewhat difficult to compute without knowing these parameters as a function of both current and temperature.
7.1.5 Radiative component estimation

Plotting $J_{th}$ for each structure as a function of temperature and extrapolating from the approximately linear low temperature region gives an estimate of the radiative current fraction at 300 K. This extrapolation relies on the assumption that at low temperatures non-radiative recombination will be minimal, which appears reasonable here based on the small threshold at low temperature, characteristic temperature and Z analysis results discussed above. In addition defect related recombination is assumed to be fairly minimal, which also seems reasonable with little evidence to suggest its presence in these devices. At temperatures where these assumptions are valid the radiative current should scale linearly with temperature based on its quadratic dependence on carrier density ($n_{th} \propto T$ and $B \propto 1/T$, therefore $J_{rad} \propto T$ from equation 2.9).

Deviation from the linear regime will take place as non-radiative recombination mechanisms are thermally activated and the fraction of injected current associated with these becomes comparable to the radiative component. If the estimated radiative current is then subtracted from the threshold current density the non-radiative current can also be estimated; this is shown for all three structures in figure 7.13 below (A, B and C top to bottom). As all analysis thus far has suggested, the non-radiative contribution to recombination decreases significantly with increasing composite well thickness, specifically $J_{non-rad}$ was found to be 84±6 %, 76±6 % and 59±6 % for A, B and C respectively. The change from A to C is certainly statistically significant, and so the trend overall fits expectation, however the change from A to B is within the limits of the uncertainty on the calculation making an analysis of the difference between these structures specifically unreliable. The
large non-radiative fractions highlight that there is a good amount of scope for optimisation and improvement with these devices through the identification and characterisation of the mechanisms responsible. These fractions give a starting point from which to expand on with further study, to firstly confirm the validity of these numbers and to de-couple the non-radiative term into its constituent components associated with individual recombination mechanisms.

![Figure 7.11: The temperature dependent threshold current density and approximate radiative and non-radiative threshold current density components. Top: Structure A. Middle: Structure B. Bottom: Structure C.](image)

7.1.6 **The Band gap as a function of temperature**

The lasing wavelength was also measured as a function of temperature for the three structures by taking spectra above threshold (110% of $J_{th}$) using an OSA and measuring the peak wavelength. The normalised spectra are shown in the right hand plot of figure 7.12 (alternate temperature points omitted for clarity) and the emission energy as a function of temperature in the left hand plot of figure 7.12. The lasing energies of all three structures were found to have fairly similar temperature dependencies of 0.41, 0.40 and 0.40 meV/$K^{-1}$ for A, B and C respectively, with a trend that fits the expected Varshni form. A consistent difference of $\sim 30$ meV (56 nm at 300 K) is observed between structures A and C (B and C consistently differ by $\sim 3$ meV or 8 nm). This is fairly consistent with the $\sim 40$ meV difference between the simulated e1 to HH1 transition energies of structures A and C, as shown in figure 7.3. The 10 meV difference between the two values is attributed to a combination of increased band filling in A compared to C resulting from the larger threshold current density and narrower well.
7.1.7 Temperature dependence conclusions

Temperature dependent analysis has given strong indications of the efficiency limiting processes being highly temperature and carrier density dependent. This highlights carrier leakage and Auger recombination as likely candidates for the dominant non-radiative recombination processes for these devices. A low $T_0$ value at ambient temperature and $Z$ value close to 3 both support the hypothesis of Auger recombination being dominant in all three structures, as does the fact that lasers operating around 1550 nm are limited by Auger recombination in a more general sense [81].

The reduction in $T_1$, $J_{th}$ and $Z$ as well as the increase in $T_0$ with increasing composite well thickness (A to B to C) all support Auger recombination and carrier leakage since both are highly dependent on band filling. The increase (decrease) in emission wavelength (energy) of 56 nm (30 meV) from A to B coinciding with a 49 % reduction in $J_{th}$ would support carrier leakage as a dominant non-radiative recombination mechanism since it is so highly dependent on the confining potential of the carriers in the QWs. Further to Auger recombination and carrier leakage the presence of optical losses is also a possibility based on the strong reduction in $T_1$ with increasing temperature.

The overall non-radiative threshold current densities were estimated based on the temperature dependence of the threshold carrier density to be 84±6 %, 76±6 % and 59±6 % for A, B and C respectively indicating that the proportion of non-radiative recombination decreases significantly from A to B to C. The application of high hydrostatic pressure dependent electroluminescence compliments this study and provides another way to investigate carrier recombination, as is now discussed.
7.2 Hydrostatic pressure dependent characterisation

7.2.1 Introduction

The application of hydrostatic pressure to a semiconductor sample will compress it equally in all directions causing a uniform reduction in lattice constant. When the pressure is removed the lattice is able to expand back to its ambient pressure equilibrium. Since the gamma point band gap scales inversely with lattice constant it can be altered reversibly using this technique. The pressure coefficients of the various bands ($\Gamma$, $X$, $L$) will all vary (and may even be negative) for different materials and so as well as varying the absolute values of the band gaps, the relative alignment of these bands will also vary. Thus, band engineering that would otherwise require the growth of various different structures may be reversibly carried out on individual devices, making this characterisation technique particularly powerful as well as time and cost effective. The details of the system used to apply the pressure and its operational procedure is detailed in section 3.4 of chapter 3.

Since the pressure dependent investigation revolves around the core principle of manipulating the band structure of the devices it is useful to have a baseline band alignment at ambient temperature and pressure in order to aid the interpretation of results. Figure 7.13 depicts the band alignment for structure C (2079) including the bulk $\Gamma$ band edges (black), confined $\Gamma$ states (solid blue), L minima (dashed blue) and X minima (dashed red). The material parameters and interpolation schemes required for the simulations were taken from Vurgaftman [102] and the simulations carried out using the nextnano software package by Dr. S. Jin within the ATI at the University of Surrey.

From the temperature dependence study, Auger recombination has been suggested to be the most prominent recombination mechanism in these devices. The Auger component may be high as a result of the SO splitting energy of 760 meV, being very close to the 786 meV of the e1 - hh1 transition of the QW. The similarity between these two energies gives a strong possibility for resonant or near resonant CHSH Auger recombination [147] in accordance with equation 2.14 in section 2.3.2 of chapter 2. If this is the case then one would expect a decrease in Auger recombination with pressure since this will cause an increase in the band gap, but have minimal impact on the SO splitting energy and thus move the CHSH process away from resonance, increasing its activation energy.

The second non-radiative recombination mechanism for which there was evidence in these devices is carrier leakage. The main parameter which determines the magnitude of the leakage current is the effective confining potential of the carriers within the QWs. For the situation of an un-pumped system at thermal equilibrium, as shown in figure 7.15 the confining potential is 215 meV, which is approximately $8k_BT$ at 300 K, meaning that even the most energetic carriers would be $\sim$100 meV short of escaping the QWs. This figure does however assume that the Fermi-level splitting will not progress beyond the e1 and hh1 levels, which would increase the probability of carriers having the energy to escape into the barriers. In addition the cladding X minima are found only a further 85 meV above the barrier $\Gamma$ minimum, opening up the possibility for leakage via the indirect valleys. The confining potential for the free holes in the valence band is slightly larger, at 263 meV.
and in addition the larger effective mass of the holes makes leakage less probable in general, thus suggesting that the dominant leakage path would be through the conduction band states.

![Figure 7.13: The calculated ambient temperature and pressure band alignment for structure C (2079) including barrier and cladding X and L minima.](image)

### 7.2.2 Electroluminescence

The LI characteristics (figure 7.14) as a function of pressure exhibit a strong increase in threshold current density for all three structures, with an approximate limit of $\sim 1750 \text{ Acm}^{-2}$. Beyond this the power supply hit its current limit of 2 A, preventing further data points from being produced. It should be noted that the pressure at which each structure reaches this threshold current density varies strongly; 232 MPa, 351 MPa and 432 MPa for A, B and C respectively. The ambient pressure thresholds impact this, giving B and A particularly less scope for increase before reaching the limit of the characterisation system. In addition, as can be seen from figure 7.15 the rate of increase of threshold with pressure also varies with $A > B > C$. The unstable nature of the LIs, particularly in the case of structure A (top plot in figure 7.14) is attributed to vibrations within the system; and the fact that the emission had to be coupled into a detector mounted externally to the pressure cell, rather than a bare fibre placed in near contact with the device.

The strong increase in threshold current density does not fit with Auger recombination alone being the dominant recombination mechanism, since as mentioned above, the CHSH mechanism should move away from resonance with increasing pressure. The presence of Auger recombination may however be masked if there is at least one other process with an opposite and significantly stronger pressure dependence. Carrier leakage could fit this profile but would depend on the absolute change in band gap as well as the rate of change. In order to confirm that this process is carrier leakage we require the pressure dependence of each band within the structure, but particularly the relative alignments of the $e_1$ level of the quantum wells and
barrier conduction band $\Gamma, X$ and $L$ minima. The similar trends in threshold current density as a function of pressure across the structures suggests that although the proportion of non-radiative recombination decreases from A to B to C, one of the proposed two mechanisms is not made completely insignificant and that both (if the hypothesis of a combination of Auger recombination and carrier leakage is correct) remain present, but with differing magnitudes and ratios of total recombination.

Figure 7.14: Pulsed electrical injection LI characteristics collected from the facet over the pressure range 0 - 432 MPa for structures A, B and C (top to bottom).
7.2.3 Determining band alignments

In order to determine the QW energy levels as a function of pressure, the pressure coefficient of the lasing emission energy is required. In addition the assumption that the valence band alignment will remain relatively unchanged with pressure is required [105]. Figure 7.16 shows this data for all three structures, indicating the offset in energy between structure A and B/C observed from the temperature dependence study and that all three have a pressure coefficient of \(0.11 \text{ meV MPa}^{-1}\) (11 meV kbar\(^{-1}\)). The near indistinguishable pressure coefficients are to be expected since the structures differ only in QW thickness and strain, with no changes in material parameters that might affect the reaction of the lattice to pressure. The data can clearly only be collected for pressures at which the devices are found to lase, however the similarities in pressure coefficient and good linear fit make extrapolation of the lasing energy to 400 MPa for A and B a reasonable approximation.
Figure 7.16: Lasing energy as a function of hydrostatic pressure at 300 K for all three structures.

The other band alignments were calculated based on the pressure coefficients of the binaries, combined according to the relevant interpolation schemes, both of which were taken from Vurgaftman [102]. Using the ambient pressure band alignments from figure 7.13 as a starting point, along with these pressure coefficients, the predicted alignments as a function of pressure can be calculated. These are plotted along with the experimental data shown in figure 7.16 in figure 7.17 below relative to the QW SO band edge. The pressure dependencies are fairly weak and so the percentage change in band gap over 400 MPa is reasonably minimal. However, the fact that the valence band alignment remains essentially unchanged means that the change in conduction band alignment can be considerable, particularly when the offsets are only a few hundreds of meV to start with. It is of interest to note that the negative pressure coefficient of the X-minimum for the cladding and barriers moves both of these band edges closer to the lasing e1 level and actually makes the barrier X minima the lowest neighboring level to the QWs when pressure is applied.
Figure 7.17: Calculated band alignments as a function of pressure using the published binary pressure coefficients and interpolation schemes to modify the band gaps shown in figure 7.15 along with the experimentally determined lasing energies.

Adjusting figure 7.13 to take into account the changes in alignment at 400 MPa gives figure 7.18, which shows the new conduction band offsets and $E_g - SO$ energy difference. The changes in band gap which appeared fairly minimal from figure 7.17 can be seen to have quite a considerable effect on the conduction band alignments. The $e_1$ to barrier $\Gamma$ offset has been reduced from 215 meV to 130 meV, nearly halving the confinement of free electrons, furthermore since the barrier $X$ minima have a negative pressure coefficient this band edge is now lower than the $\Gamma$ minimum, making the barrier indirect and creating an alternative potential leakage path with a confinement energy of only 100 meV. These two changes would be consistent with carrier leakage being the cause of the rapid increase in threshold current density with pressure but do not identify which of the two represents the leakage path.

The cladding $X$ minimum is now found only 55 meV above the barrier $\Gamma$ minimum but remains 85 meV above the barrier $X$ minimum. Therefore, leakage from the $e_1$ to barrier $\Gamma$ minimum to the cladding $X$ minimum would be expected to increase. Leakage from the barrier $X$-minimum to the cladding $X$-minimum, however should remain reasonable constant since the alignment has not changed, assuming the increased bias required to reach threshold at 400 MPa is not taken into account.

The energy of the $e_1 - hh1$ transition has been raised to 846 meV, increasing the separation from the SO band from 16 meV at ambient pressure to 86 meV at 400 MPa, a change of $\sim$540%, which from equation 2.14 increases the CHSH Auger recombination activation energy by the same factor. Thus, the CHSH Auger recombination fraction would be expected to fall off drastically with pressure, which again would imply a
reduction in threshold current with pressure if this were the only significant recombination mechanism present, further supporting the hypothesis of a combination of carrier leakage and Auger recombination being the key efficiency limiting processes. Other Auger recombination processes such as CHCC that may also be taking place would be effected to a lesser extent relatively from the increase in bandgap, but this still would not explain the strong increase in $J_{th}$.

Figure 7.18: The calculated ambient temperature band alignment for structure C (2079) including barrier and cladding X and L minima for ambient pressure (left) and 4 kbar hydrostatic pressure (right).

If the threshold current density vs pressure plot is normalised then the trends in carrier leakage can be plotted along side using the relation $J_{leak} = J_0 \exp \left[ \left( \frac{-dE_b}{dP} \right) P/k_b T \right]$ where $J_0 = 1$ for the case of normalised data. The parameter $\frac{dE_b}{dP}$ is the net change in band alignment as a function of pressure between the $e_1$ state and the proposed leakage state, in this case the barrier X minimum, here the $e_1$ energy increases at $\sim 11$ meV/kbar$^{-1}$ and the X-minimum decreases at $\sim 1.8$ meV/kbar$^{-1}$, giving a net value of $\frac{dE_b}{dP} = 11 - 1.8 = 12.8$ meV/kbar$^{-1}$. The theoretical case of 100% radiative recombination relies on the approximation $J_{rad} \propto E_g^2$, giving a simple relationship between $J_{rad}$ and pressure based on the pressure coefficient of the lasing energy as follows:

The change in bandgap as a function of pressure is simply the value at ambient pressure plus the product of the pressure coefficient and applied hydrostatic pressure:

$$E_g(P) = E_g(0) + \frac{dE}{dP} \cdot P. \tag{7.1}$$

Combining equation 7.1 with the approximation that $J_{rad} \propto E_g^2$ gives equation 7.2,

$$J_{rad} \propto \left( E_g(0) + \frac{dE}{dP} \cdot P \right)^2 \tag{7.2}$$
relating the radiative current density to the applied pressure. These data sets are plotted together in figure 7.21 below and a combination of carrier leakage and radiative recombination is used as a fit to the normalised threshold current density data. The fit simply combines the radiative and leakage currents together using linear interpolation and varies the fraction of each to give the best fit, the radiative fractions are found to be 75%, 68% and 56% for A, B and C respectively. Although this fit is found to be reasonable it is clearly unphysical since from both the pressure and temperature dependence studies it is shown that C > B > A in terms of efficiency and performance. The radiative fractions for A, B and C were approximated to be 16%, 24% and 41% respectively, not matching these fits at all. Therefore we can say that for the fit to make physical sense a third term is required, which based on the previous data is likely Auger recombination. The trends in performance of the devices vs pressure thus far have only indicated the presence of carrier leakage and so this result helps in matching the conclusions of the temperature dependence to the pressure dependence study.

Here the monomolecular recombination component has been assumed to be negligible based a lack of evidence of its presence from this study and the fact that these samples are grown on a GaSb, rather than a Si substrate using mature, well understood growth procedures.

Figure 7.19: Normalised current density, theoretical 100% leakage (e1 to X minimum path) and theoretical 100% radiative recombination for all three structures as a function of pressure

In addition to the lowest energy leakage path from the e1 level to barrier X-minimum there is also the possibility of leakage into the barrier Γ minimum to be considered. The Γ-minimum of the barriers has a pressure coefficient of 5.5 $meV/kbar^{-1}$, giving a net value for the coefficient of the leakage path is $\frac{dE_l}{dP} = 11 - 5.5 = 5.5$ $meV/kbar^{-1}$, far less than the 12.8 $meV/kbar^{-1}$ of the e1 to X-minimum path.

Thus the leakage current via this path would be expected to increase at a reduced rate, in accordance
with the band alignments found from figure 7.18. The theoretical normalised 100% leakage limit data is plotted alongside the normalised threshold current data in figure 7.19 and it can be immediately seen that this process cannot be responsible for the strong pressure coefficient of $J_{th}$. If this leakage path is assumed to be responsible for the bulk of the carrier leakage then there is no other mechanism with a great enough pressure coefficient to bring that of $J_{th}$ up to its measured value.

Having already seen from the analysis of figure 7.19 that carrier leakage and radiative recombination alone are not enough to explain the pressure dependent electroluminescence results the calculated trend for Auger recombination has also been added to 7.19. The trend in normalised Auger component is found by taking the pressure coefficient of the lasing energy to SO splitting difference, which given that the valence bands remain effectively static is $11 \text{meV kbar}^{-1}$, and applying this to equations 2.13 and 2.14 of chapter 2. It is noted that in addition to the $E_g$ - SO splitting the Auger activation energy is also dependent on the effective masses of the bands, which have their own pressure dependencies. Furthermore the Auger current is proportional to $n_{th}^3$, which will scale with pressure also as has been shown by the large change in threshold current density. However, it is not easily predictable since a lot of the carriers may simply be passing the QWs entirely via carrier leakage paths. Due to the complexity of determining these parameters (particularly with unreliable spontaneous emission data) the approximate normalised Auger current density is found by using the results of previous studies into GaInAs [135] and modifying according to the $E_g$ - SO splitting. This gives the trend plotted in figure 7.19 as an estimate of how the normalised threshold current density would vary with pressure for a theoretical 100% Auger recombination dominated device. As was expected, this trend decreases with pressure and so cannot be combined with the leakage path via the $\Gamma$-minimum of the barriers but would make physical sense as an additional mechanism if leakage is assumed via the X-minimum.

The non-radiative component calculated at threshold from figure 7.11 above is plotted in figure 7.20, along with the trends for radiative, Auger recombination and carrier leakage through the barrier X-minima. Doing so gives an indication of the combination of carrier leakage and Auger recombination that would be required to fit the experimental data. It is found that all three devices have a large carrier leakage component based on the large, positive gradient and that the component for leakage on a normalised scale is largest for C, then B and lowest in A. Since this is on a normalised scale it does not take into account that the threshold current density for C is less than half that of A at ambient pressure and temperature, and so the magnitude of the leakage current is actually likely less in absolute terms but proportionately larger in C.
Combining these normalised trends it is possible to make a fitting function to the normalised threshold current density according to equation 7.1:

\[ J_{th} = X \times J_{rad} + (1 - X)[Y \times J_{leak} + (1 - Y) \times J_{Auger}] \] (7.3)

The parameter X defines the fraction of \( J_{th} \) associated with radiative recombination, leaving (1 - X) as the non-radiative fraction, with Y and (1 - Y) the leakage and Auger fractions respectively. For the radiative fraction, X the approximations made during the temperature dependence study, from figure 7.11 are 16±6%, 24±6% and 41±6% for A, B and C respectively. With this parameter known Y is varied until the best fit is achieved, giving the approximate leakage and Auger fractions at ambient pressure that would lead to this trend in threshold current density with pressure.

These functions are plotted in figure 7.21 and found to give reasonable fits for radiative, leakage and Auger fractions of \( J_{rad}=16\% \), \( J_{leak}=28\% \) and \( J_{Aug}=56\% \) for A, \( J_{rad}=24\% \), \( J_{leak}=33\% \) and \( J_{Aug}=43\% \) for B and \( J_{rad}=41\% \), \( J_{leak}=39\% \) and \( J_{Aug}=20\% \) for C (an error of approximately ±6% on each). It can be seen that the Auger fraction decreases considerably, from 56% to 20% from A to C, which would be consistent with the large difference in ambient temperature threshold current density for the two devices.

The leakage fraction changes to a lesser degree, increasing from 28% to 39% (A to C), which is somewhat counter-intuitive since firstly the band filling should decrease with increasing gain material volume and modal gain. Secondly, as can be seen from figure 7.17 the unfilled e1 level decreases in energy from A to C, again leading to the expectation of decreased leakage as a result of an increasing confining potential, \( E_B \). However,
a 36% drop in Auger recombination has led to only an 11% increase in radiative recombination, and so it may be that although the leakage coefficient is dropping, a proportion of the carriers that previously recombined via Auger recombination are now escaping the wells via a leakage path. If this were the case the leakage fraction would decrease less (or even increase) and the radiative fraction would increase less than expected, as is observed here. It should be noted of course that these are just fractions and not absolute values and so a greater leakage fraction does not necessarily mean a greater absolute leakage current.

![Graph showing normalized threshold current densities and fitting functions for structures A, B, and C as a function of pressure.]

Taking the fractions at ambient threshold and multiplying by the ratio of the normalised current density components to the threshold current density for each pressure point gives the trend in fractional recombination mechanisms for each structure, as plotted in figure 7.22. The three structures (A to C, top to bottom) all exhibit the same general trend, with leakage starting out as the intermediate recombination mechanism and rapidly becoming the major one, and in B and C even accounting for \(~80\%\) of injected current at the highest pressure.

The Auger component also decreases strongly, as is to be expected as a result of the increased difference between the lasing energy and the SO band energy, but can be seen to be a far more impactful recombination mechanism in A and B than C. The increased dispersion of carrier energies in A and B will be a major factor in increasing the probability of Auger recombination, as will the lower modal gain, increasing the threshold current density, which in turn increases the fraction of Auger recombination since \(J_{\text{Aug}} \propto n_{th}^3\), whilst \(J_{\text{rad}} \propto n_{th}^2\). The reduction in radiative fraction is simply a result of such a strong increase of leakage current.
Finally, by taking the normalised fractional contributions of each recombination mechanism and multiplying by the experimentally determined threshold current densities, the absolute values of current density associated with each recombination mechanism are found as a function of pressure and plotted in figure 7.23. Plotting on an absolute scale aids in this case since the threshold current densities at any given pressure vary so strongly, making it appear that the vastly superior performing device C has a greater leakage current than A if the normalised fractions are used only. Here it can be seen that C (crossed data) has the greatest radiative current at all pressures, which is particularly impressive considering its total threshold current density is approximately half that if A at ambient pressure. In addition it consistently exhibits both the lowest Auger recombination current and leakage current even though it has the largest leakage fraction.

Of particular interest is the Auger recombination current density of A (solid red data), this starts at $\sim 560 \text{ Acm}^{-2}$ (Ambient pressure), falling to $\sim 500 \text{ Acm}^{-2}$ at 225 MPa, the ambient pressure Auger current density alone is larger than the $J_{th}$ of structure C at ambient pressure, 471 $\text{ Acm}^{-2}$. As a function of pressure, all devices end up limited by carrier leakage, as one would expect based on the decrease in offset to the barrier X-minimum with pressure and increasing lasing energy to SO energy separation decreasing CHSH Auger recombination, as well as a general decrease in Auger recombination more generally in accordance with the findings of [135].

Figure 7.22: Fractional normalised threshold components as a function of pressure for A, B and C (top to bottom).
Both the fractional and absolute values of the recombination mechanism components of $J_{th}$ are summarised below in Table 6 for ease of comparison along with the total $J_{th}$ values measured at ambient pressure, $T_0$, $Z$ and $E_{lase}$ at ambient temperature and pressure. It is clear that at ambient pressure, where the devices will actually operate Auger recombination is dominant for both A and B at 352% and 179% of $J_{rad}$ respectively, all be it to a far lesser extent in B than A, and falls off considerably to only 48% of $J_{rad}$ for structure C. It is interesting to note that the Auger component of structure A is so large, since from figure 7.12 it can be seen that the lasing wavelength is 56 nm less than in C, corresponding to an increase in $e_1 - hh_1$ energy of 29 meV. This moves the CHSH processes further from resonance than either B or C, for which Auger recombination is found to be less. This result highlights the importance of maximising the gain volume, strain and modal gain in order to reduce the threshold carrier density. The implication is that these factors increase the threshold carrier density to such an extent that cubic dependence of the Auger current density on $n_{th}$ increases the Auger fraction (from a combination of multiple Auger processes) enough to counteract the lack of resonance (for the CHSH processes specifically) in this case.

Since the CHSH Auger recombination processes is sensitive to the $e_1$-$hh_1$ to SO energy separation increasing this lead to a further reduction in the CHSH Auger recombination process. This is however a difficult if one wishes to maintain an emission energy close to 1550 nm, the SO band has however been shown to be strongly increased by the additions of small fractions of bismuth [69] in both GaAs and InAs bases material systems. This could be a potential approach to increasing the spin-orbit splitting to the point of exceeding the
energy of the e1-hh1 transition, thus making the CHSH process effectively impossible (CHCC is unaffected however).

Leakage for the three structures is the minor of the two non-radiative mechanisms for A and B, at 174% and 140% of \(J_{rad}\) respectively but becomes the major non-radiative recombination mechanism for C with a value of 96% of \(J_{rad}\). This does however show that of the three mechanisms radiative recombination is actually the major mechanism in the case of C, unlike A and B indicating that this structure, with some optimisations could be highly efficient.

To reduce carrier leakage an increase in the activation energy of the leakage path would be required. If the e1-hh1 energy is to be maintained in order to achieve lasing at 1550 nm then the barrier energy would have to be increased. Calculating a Fermi-Dirac probability distribution at 300K and converting to a normalised cumulative probability the difference in energy between 39% (from leakage fraction of injected carriers) and 20%, 10%, 5% and 1% of carriers having the energy to escape the QWs is approximately 20 meV, 38 meV, 57 meV and 99 meV respectively. Based on these rough numbers the carrier leakage fraction could be halved by increasing the barrier band gap by as little as 20 meV, reduced by a factor of 4 by an increase of \(\sim40\) meV or almost completely eliminated by an increase of \(\sim100\) meV. Thus if the composition of the structure is required to be close to its current values for other reasons such as lattice and thermal expansion coefficient matching or refractive index minimisation to achieve good optical confinement, then a fairly minimal increase of 20 meV could be chosen in order to halve carrier leakage and reduce the threshold current density by \(\sim20\)%. 

<table>
<thead>
<tr>
<th></th>
<th>(J_{th}^{rad}) (%)</th>
<th>(J_{Aug}^{rad}) (%)</th>
<th>(J_{leak}^{rad}) (%)</th>
<th>(J_{rad}^{(Acm^{-2})})</th>
<th>(J_{Aug}^{(Acm^{-2})})</th>
<th>(J_{leak}^{(Acm^{-2})})</th>
<th>(T_0(K))</th>
<th>(Z)</th>
<th>(E_{lase}[meV])</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>16 ± 6</td>
<td>56 ± 6</td>
<td>28 ± 6</td>
<td>159 ± 60</td>
<td>559 ± 60</td>
<td>276 ± 60</td>
<td>995</td>
<td>63 ± 6</td>
<td>2.91 ± 0.08</td>
</tr>
<tr>
<td>B</td>
<td>24 ± 6</td>
<td>43 ± 6</td>
<td>33 ± 6</td>
<td>150 ± 38</td>
<td>268 ± 38</td>
<td>210 ± 38</td>
<td>628</td>
<td>68 ± 6</td>
<td>2.85 ± 0.08</td>
</tr>
<tr>
<td>C</td>
<td>41 ± 6</td>
<td>20 ± 6</td>
<td>39 ± 6</td>
<td>195 ± 29</td>
<td>94 ± 29</td>
<td>187 ± 29</td>
<td>476</td>
<td>92 ± 6</td>
<td>2.83 ± 0.08</td>
</tr>
</tbody>
</table>

Table 7.1: The fractional and absolute calculated recombination components of \(J_{th}\), \(T_0\), \(Z\) and \(E_{lase}\) at ambient temperature and pressure for A, B and C.

### 7.2.4 Pressure dependence conclusions

The results of the pressure dependence study align well with those of the temperature dependence, highlighting a combination of Auger recombination and carrier leakage as the dominant sources of loss in these devices. From the simulated band alignment diagram it was found that the lasing energy to SO band difference was only 16 meV at ambient pressure (noting that \(k_BT\) is \(\sim26\) meV), causing to CHSH Auger recombination mechanism to be close to resonance. In addition the negative pressure coefficient of the barrier X-minimum opened the possibility for a pressure activated leakage path that would limit performance at increased pressure. The ambient pressure confinement for an unpumped device between the QW e1 state and the barrier \(\Gamma\) (or X, they have the same energy at ambient pressure) minimum is 215 meV, which at 300 K should be sufficient to confine the majority of carriers. However, at threshold the band filling may extend considerably beyond the e1 level as a result of both pumping and the asymmetrical nature of the quasi Fermi-level splitting in a system with dissimilar electron and hole effective masses.
Using the published pressure coefficients of the binaries and combining by interpolation the pressure coefficients of the bands for the barriers and cladding layers were estimated at increased pressure levels. In combination to the pressure coefficient of the lasing transition, determined experimentally from the lasing energy as a function of pressure the band alignments as a function of pressure were calculated. It was found that the lasing energy to SO band splitting increased from 16 to 86 meV, leading to a prediction of reduced Auger recombination with pressure. The confining potential of the QWs was found to decrease from 215 meV to 100 meV if the leakage path is assumed to be via the barrier X-minimum (the lowest energy barrier state), indicating that carrier leakage should increase significantly with pressure. By estimating the trends in radiative recombination, Auger recombination and carrier leakage it was possible to fit a function of these three mechanisms to the normalised threshold current densities and use the coefficients of the fit to estimate the fractions of each at threshold. To reduce the number of unknowns the radiative fraction was taken from the temperature dependence data, leaving the non-radiative fraction to be split between Auger recombination and carrier leakage to achieve the best fit.

The fractions found for each mechanism (summarised in table 6 above) supported the implications of the electroluminescence results that both Auger recombination and carrier leakage accounted for significant proportions of injected current and that both mechanisms were reduced strongly from A to B to C as the composite well thickness increased, raising both the gain volume and modal gain. Structure C, evidently superior from the LIs alone is found to have radiative recombination as the dominant recombination mechanism, closely followed by carrier leakage, with much less, but still significant contributions from Auger recombination. These pressure dependence results highlight how sensitive to fairly small changes in band alignment both Auger recombination and carrier leakage are and that with even minor alterations large increases in efficiency and eventually a commercially viable laser could be achieved.

7.3 Conclusions

Temperature dependent characterisation indicated the strong dependence of both $\eta_d$ and $J_{th}$ on both temperature and carrier density. The low $T_0$ values of $63 \pm 6$ K, $68 \pm 6$ K and $92 \pm 6$ K at ambient temperature, $Z$ value close to 3, reduction in $T_1$, $J_{th}$ and $Z$ as well as the increase in $T_0$ with increasing composite well thickness (A to B to C) support a combination of Auger recombination and carrier leakage as the dominant source of loss in these devices. On top of the non-radiative recombination mechanisms optical losses are also a possibility based on the strong reduction in $T_1$ with temperature. The radiative fractions of the threshold current density were approximated by extrapolation of the low temperature threshold current density (where no radiative recombination should be minimal) to 300 K and is found to be $\sim 16\pm 6\%$, $24\pm 6\%$ and $41\pm 6\%$ for A, B and C respectively. These values fit with the hypothesis of an Auger recombination and carrier leakage limited material system that is greatly improved by the increase in gain volume and modal gain with increasing composite well thickness from A to B to C.

Expanding on the results of the temperature dependence study the application of hydrostatic pressure was used to vary the band alignment reversibly and in doing so test the hypothesis of an Auger recombination and
carrier leakage dominated device and quantify the contributions of each. The similarity between the e1-hh1 energy and SO band energy identified by simulation of the band alignments using the nextnano software (see figure 7.18) would cause the CHSH Auger processes to be close to resonance at ambient pressure. The strong increase in threshold current density with pressure fitted with a carrier leakage dominated system. Fitting the trend in normalised threshold current density with pressure required both carrier leakage via the barrier X-minima and Auger recombination in addition to radiative recombination to provide a good fit and make physical sense simultaneously. Taking the radiative fractions to be those determined from the temperature dependence study the Auger and leakage fractions were found from the fitting parameters and are summarised in table 6 above. These values highlighted that Auger recombination is significant in all three structures, but to a much greater extent in B and particularly A. Carrier leakage is then left as the dominant non-radiative mechanism in the case of C but remains a major source of loss in both A and B, just to a lesser extent than Auger recombination.

This study has identified and quantified the efficiency limiting processes for the GaSb based lasers produced by the group of E. Tournie at Montpellier University to be Auger recombination and carrier leakage. The use of a composite quantum well to increase the gain material volume as well as modal gain and strain clearly has huge benefits in reducing the threshold current density and increasing the slope efficiency. These improvements are attributed to an increase in confinement for electrons within the quantum well as a result of reduced band filling and threshold carrier density. The carriers are spread over a larger volume without reducing carrier confinement or strain compared to the narrower wells, thus maintaining the desired emission energy and gain. Furthermore the optical confinement factor and therefore modal gain is increased by both the larger pumped gain volume but also the increased overlap of the optical mode with the active region, maximising the rate of stimulated emission and reducing re-absorption and optical losses. Although Auger recombination is reduced as a result of a decreased dispersion of carrier energies and reduced threshold carrier density it remains a considerable source of loss, accounting for approximately 20% of injected current at ambient temperature and pressure. In the case of structure C, which would presumably be the basis for future developments, carrier leakage is estimated to account for ~39% of recombination at ambient temperature and pressure. Attempting to reduce carrier leakage by increasing the barrier band gap by as little as 20 meV, as discussed with reference to table 6 would be fairly simple by altering the composition towards that of the SCH layers. The lower $J_{th}$ achieved in doing so would then lead to a reduced , and thus also lower the Auger recombination rate, as would increasing the the e1-HH1 to SO band energy difference (for CHSH specifically), potentially by the addition of a small percentage of Bi [69].

These devices show great promise for the realisation of efficient, high crystalline quality Si compatible lasers on CMOS compliant [100] orientated substrates. The optimisation of these structures to further reduce non-radiative recombination would produce lasers with impressive threshold current densities and efficiencies. Thus when transferring over to Si substrates the additional issues associated with the incorporation of large strain will be simpler to identify and decouple from those inherent to the material system and thus address appropriately.
8 Conclusions and Future Work

The main intended goal of this thesis was to identify and quantify the extent of the physical processes which limit the efficiency of three of the four leading approaches to achieving optoelectronic integration on Si with III/Vs. Further to these findings, optimising and improving upon the devices studied here through structural changes, band engineering and fabrication process changes was carried out where possible in order to maximise performance.

In order to meet these goals the general approach was to initially characterise complete devices using temperature and pressure dependent electroluminescence. Since electroluminescence gives easy access to important characterising parameters such as the slope efficiency and threshold current density an initial idea of performance can be quickly yielded, particularly through the use of the probe station system. As is detailed in sections 2.2 and 2.3 the recombination mechanisms which determine the level of performance of the devices have well defined, discernible trends with temperature and pressure. Thus, electroluminescence as a function of these two parameters is a robust mode of investigation that yields maximum information from each experiment. By collecting both stimulated emission from the facets and pure spontaneous emission from the substrate in both an integrated intensity form as well as spectra with the use of an OSA additional data on alternative loss processes such as IVBA or detuning effects are also collected from this single type of experiment.

Further specific methods of study were also applied as required for each unique type of device to fit both the physical properties and performance as well as structural choices. For example, when the GaNAsP based devices ceased to lase (preventing the continuation of the versatile electroluminescence studies), simulation of the waveguiding properties of the lasers and a processing optimisation study were carried out.

The main challenges that were faced overall related to fairly specific demands on the design and size of devices that can be accommodated easily within the closed cycle cryostat or high pressure systems. The large bars of tightly packed devices on silicon substrates of DFB waveguide hybrid lasers made direct fitment or cleaving for use with these systems close to impossible, hence the limited initial studies. The lack of micromanipulator probes in these systems also led to contacting issues on closely packed devices or those with small contacts, with both of these issues prominent in the FP cavity hybrid devices and to a lesser extent with the GaNAsP based lasers. With a small set of fragile samples a custom sample holder and contacting system was designed for the FP cavity hybrid lasers, which allowed use with the cryostat although with much repartition and additional time investment required to achieve a full set of data. The tighter constraints of the pressure system made this more or less impossible without access to a diamond saw or laser cutting system or perhaps a large enough set of samples to account for damaged devices. If this investigation were to continue individual diced devices would be the most suitable choice for use with these systems so that gold wire bonding and lasers headers could be used to achieve both good electrical contact and mechanically stable mounting.

This approach in general yielded favourable results, with the dominant sources of loss and inefficiency
determined for all three approaches studied here. In addition, the recombination mechanism fractions for both the GaNAsP based lasers and GaSb based lasers were determined, with an estimate of the fraction of SRH recombination found for the hybrid lasers. The effect on these values from various structural changes (number of quantum wells, presence of defect blocking layers, surface orientation and composite well width) was also quantified, giving clear information not only on the performance changes yielded by these changes, but also the effect on the underlying physical processes. Thus, further concepts for improvement can (and have been) developed based on an informed position, knowing not only how changes effect efficiency, but critically, why.

Bringing together the results and conclusions of each individual study presented in this thesis allows broader statements to be made on the field of III/V integration on Si for optoelectronic integration. A quick summary of the findings for the AlGaInAs-hybrid evanescent lasers, GaNAsP based monolithic lasers and GaSb based monolithic lasers are first given for comparative purposes (structure specific results are omitted at this stage). Detailed conclusions on the studies of each specific type of device can be found in the final section of each relevant chapter or section.

The AlGaInAs-hybrid evanescent lasers operated up to 50 °C at around 1600 nm, with a room temperature threshold current density of 0.5 to 0.8 kAcm⁻² per QW, single facet output power of ~ 10 mW. Although the threshold current density is 5 to 8 times the typical value of commercial 1550 nm communications lasers the other parameters otherwise seem reasonable on face value. This approach has seen fast development as a result of the use of a well understood material system, grown defect free on an InP, as oppose to Si substrate and integrated later. Thus, defects that would usually be formed primarily during growth of III/Vs on Si are avoided.

However, low yields and highly variable performance (see figure 4.4) suggest that the bonding process may not yet have been perfected. In addition, without the use of a defect blocking layer the SRH recombination fraction is as much as 39% at room temperature, dropping to a not inconsiderable 22% even with several periods of SSL above the bonding region. Threshold Z values in excess of 3, low $T_0$ and $T_1$ values of at least 91±9 K and 112±12 K (relative to the ambient temperature of 300 K) and conduction and valence band carrier confinement energies ($E_B$) of ~22 and ~23 meV respectively suggest carrier leakage dominates recombination in addition to SRH recombination. This is further exacerbated by the observation of a lack of pinning and a tendency for internal heat build up as a result of the insulating SiO₂ layers. The internal efficiency and internal optical loss were determined to be 69±4 % and 19±2 cm⁻¹ at ambient temperature respectively. Injection issues, likely related to bonding defects and optical losses, which would be expected to be IVBA from other studies of similar materials [141] [121] are also then limiting factors in these devices.

The lasers developed based on the dilute nitride active region GaNAsP monolithically grown on Si have the advantageous property of not requiring time consuming and costly additional steps, as is the case with hybrid integration. In addition other factors such as bonding defects and inefficient heatsinking are eliminated inherently. The development of a new material system however bring separate issues to the use of a proven one such as AlGaInAs/InP. GaNAsP as well as BGaP and BGaAsP have not been extensively studied causing
additional work to be required to optimise growth and design. Lasing has been observed up to 165 K with
a large threshold current density of 1.6 \( kAcm^{-2} \) per QW and the operating wavelength is only \( \sim 900 \) nm,
far from either the minimum loss or zero dispersion wavelengths of Si waveguides. Across all devices SRH
recombination has been prominent, accounting for between approximately 50\% and 65\% of recombination
suggesting that the growth of this material system is far from optimised. A sensitivity to Si surface orientation
may also cause additional issues as miscuts as small as 0.1\° have been found to cause a favouring of single
steps, leading to increased APD and threading dislocation generation. In addition evidence for the diffusion
of nitrogen from the active region has been observed to varying degrees (more so in past devices [34]) in
all devices, potentially leading to additional thermally activated defects and leakage paths. Even on GaP
substrates although lasing at room temperature was achieved, the threshold current density was 4 \( kAcm^{-2} \)
per QW, with non-radiative recombination accounting for in excess of 90\% of injected current.

The composite QW GaInSb/GaSb lasers have been found to lase at room temperature with a threshold
current of \( \sim 0.16 \) \( kAcm^{-2} \) per QW, only 60\% greater than typical commercial InGaAs lasers. These devices
operate at 1550 nm and were found to have a negligible SRH recombination and radiative, Auger and leakage
components of \( 41\% \), \( 20\% \) and \( 39\% \), supported by a \( T0 \) value of 92\% and a \( Z \) value of 2.83.
A certain level of Auger recombination is expected at this wavelength but could be minimised by reducing
the spin-orbit splitting to prevent a close to resonant CHSH mechanism. Based on band alignments, an
increase in barrier bandgap of only \( \sim 20 \) meV would be sufficient to halve the carrier leakage fraction, giving
the possible of considerable performance improvements for minimal compositional changes. These results
are of course for devices on a GaSb substrate, on Si threshold current densities of \( \sim 0.3 \) \( kAcm^{-2} \) per QW are
reported for electrical injection lasing at room temperature with single facet output power of \( 30 \) mW.

These three approaches each have their own individual limitations and specific procedures will be required
to develop each further. However, since all three are competing for the same goal they should also be compared
to each other in order to see which are most developed and have most potential or limitations. The hybrid
approach has exhibited a lot of progress, making it appear very promising, however limitations relating
to the bonding mechanism are inherent to this system and are the cause of almost all efficiency limiting
processes and attributes of these lasers. Development of the bonding process would therefore have to be the
primary approach to optimisation. The group responsible for these devices have however moved towards study
of quantum dot based approaches to optoelectronic integration, suggesting that a plateau may have been
reached in development. As such, at this stage although the devices show reasonable performance relative to
competing approaches they would appear to have the least potential for realisation of a commercially viable
product.

In comparison the dilute nitride GaNAsP/BGaP/Si lasers have the fundamental benefit of monolithic
integration but have so far exhibited the worst overall performance in terms of lasing characteristics. Poor
performance even on non Si substrates shows that considerable work is still required to reach the performance
levels of competing lasers. Unlike the hybrid approach however, this system theoretically has fewer fundamental
limits and with appropriate growth conditions, band engineering and fabrication should be able to
improve greatly. The lack of lasing devices, sensitivity to Si surface orientation, potential nitrogen diffusion and low emission wavelength are however problematic to achieving further progress. Thus, the conclusion is that relative to competitors the dilute nitride approach has significant promise, but requires considerable further research to achieve its potential. The defining factor in the success of this approach then, is whether or not the development progresses at a sufficient pace relative to competitors, rather than a fundamental problem.

Of the three, the GaInSb/GaSb lasers have achieved both the greatest performance and lasing characteristics as well as lacking any large fundamental issues that must be overcome. The limiting factors of Auger recombination and carrier leakage could both potentially be reduced in dominance by suggest band engineering approaches, leading to further gains. Requiring a reduction of only around 60% in room temperature threshold current density per QW to be comparable to commercial grade devices the lasers on GaSb substrates could soon be competitive. A study of these lasers on a Si substrate, where threshold current densities are greater would of course be required for a fair comparison. It is therefore concluded from this thesis that overall these lasers show a lot of potential, being monolithic and lacking any apparent additional fundamental issues specific to this approach that others do not, as well as exhibiting the best room temperature performance.

8.1 Future work

Based on the above conclusions several areas of interest suitable for further study have been identified.

A hydrostatic pressure dependence study was not carried out on the hybrid AlGaInAs silicon-evanescent lasers since the devices received were too large to be mounted in the pressure cell and too thick and densely packed to be cleaved by hand. Thus, if individual devices could be sourced a hydrostatic pressure dependence study could be carried out to compliment the temperature dependence results. This could allow quantification of the non-radiative recombination mechanisms and potentially yield greater details of the source of the non-pinning behaviour. The degree to which Auger recombination in particular is taking place may be indicated by the pressure coefficient, something that temperature dependence alone could not determine due to the strong carrier leakage.

The change of Si substrate is thought to be related to the drop in performance of the GaNAsP based lasers post #16195. As was seen in chapter 5, Si surface orientation has a strong impact on defect density for even minor changes. Thus a repeat of this study, for a greater range of miscut angles on devices produced using the new substrate may be of value in determining if the substrate is the cause of the performance drop. It may be that a minor change in surface morphology causes the miscut that was previously found to be optimal to no longer be so, leading to increased APD and therefore defect density, preventing the device from lasing.

Future GaNAsP device batches could also include variants with the optimised waveguiding structure, particularly any that are found to lase to determine if this has as significant an impact in practice as it does theoretically. In addition carrying out two processing runs in parallel using the original and optimised
methodology could give a more accurate representation of the impact this has, again particularly if a structure which lases is produced.

The GaInSb/GaSb lasers have been found to operate at room temperature and with the largest composite well width are limited mainly by carrier leakage, with a secondary source of loss from Auger recombination. As was found in chapter 7 increasing the carrier confinement by only $\sim 20$ meV may as much as halve the carrier leakage fraction. Thus a structure with increased barrier bandgap is recommended to be grown, on which a follow up investigation could be carried out to confirm the expected drop in carrier leakage. The addition of Bi to increase the spin-orbit splitting may also be a potential method of reducing the not insignificant Auger recombination in these devices. This would again require a repeat of the temperature and pressure dependence studies.

Following this the next logical investigation would be of GaInSb/GaSb/Si lasers to determine the effect efficiency limiting mechanisms of growth on Si. With the above studies of devices on GaSb substrates carried out a good understanding of the existing sources of loss should have been gained. Therefore the identification of those directly related to silicon should be more easily determined.

Appendices

Appendix A

Figure 8.1 depicts the way an approximation of the threshold defect current can be extracted from the $\ln(I)$ vs $\ln(\text{SE}^{1/2})$ plot, the gradient of which is the Z parameter at a given injection current. A region at low current where the gradient (Z parameter) is approximately equal to one is found and extrapolated to the point on the x-axis aligned with the data at threshold current (marked by the red dotted line). From here the log of the approximate threshold defect current is read off from the y-axis.
Figure 8.1: The ln(I) vs ln(SE^{1/2}) plot over the temperature range 220 - 300 K for the hybrid AlGaInAs silicon-evanescent lasers. Extrapolation from the low injection Z≈1 regime (I ∝ n) to threshold gives the approximate defect recombination fraction.

Appendix B

The confinement of photons within a waveguide is somewhat analogous to the confinement of electrons within a potential well and the mathematics is therefore somewhat similar. As such, just is the case with the finite well potential a confined mode will extend to some extent into the surrounding material even if TIR takes place. The quantification of this overlap, known as the evanescent field/mode, will be required to calculate the spreading of the optical mode across the structure in a simulation. The photons propagating along the axis of the cavity (x) are described by a plane wave equation (8.1),

$$E = E_y(x)exp(i(βz - ωt))$$

(8.1)

where it is assumed that the mode has no dependence on the transverse direction (y). For a mode to be considered guided the field must be oscillatory within the core of the waveguide and exponentially decaying within the cladding, as is shown in equation 8.2 below. Here $x$ and $γ$ are the transverse wave vectors of the confined mode within the core and cladding regions and $h$ is the thickness of the core region, measured from -h/2 to +h/2. Again, as is the case for quantum confinement the first and second derivatives (here the electric ($E$) and magnetic ($H$) fields are required to be continuous across the boundary, i.e. there should be no discontinuities in the optical field:
In order to meet these boundary conditions we find the following requirements on the constants:

\[ C_0 = C_1 \cos \left( \frac{k h}{2} \right) \]  

(8.3)

and

\[ \frac{\gamma}{\mu_0} C_0 = \frac{\kappa}{\mu_0} C_1 \sin \left( \frac{k h}{2} \right) \cdot \]  

(8.4)

Equations 8.3 and 8.4 include both unknowns, \( C_0 \) and \( C_1 \), and so can be re-written in the form of a 2x2 matrix,

\[
\begin{bmatrix}
1 & -\cos \frac{k h}{2} \\
\frac{\gamma}{\mu_0} & -\frac{\kappa}{\mu_0} \sin \frac{k h}{2}
\end{bmatrix} \begin{bmatrix}
C_0 \\
C_1
\end{bmatrix} = \begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

(8.5)

in order to find a solution. Matrix 8.5 only has a solution for the case of the determinant being equal to zero. By asserting this condition the characteristic equation, 8.6 is yielded:

\[ \gamma = \kappa \tan \left( \frac{k h}{2} \right) \cdot \]

(8.6)

The periodic nature of the function indicates the potential for multiple solutions (pairs of \( \gamma \) and \( \kappa \)), which in practice means multiple TE modes of even symmetry may be confined within the waveguide. Note this solution is only valid for even modes; 0, 2, 4..., however for the purposes of this study we require only the solution for the fundamental mode and so this is not an issue.

The V-parameter, given by equation 8.7,

\[ V = \left( \frac{k_0 h}{2} \right) (n_1^2 - n_2^2)^{1/2} \]  

(8.7)

can be considered to be related to the degree of confinement of the mode within the core region. The guided modes are then determined as the cross over points of function 8.6 and the square of 8.7, from which the propagation constant \( \beta \) can be determined. Since the modes decay exponentially into the cladding they theoretically have infinite extent and so a consistent measure for the extent of the optical field is required. The mode field diameter (MFD) is defined as twice the distance from the centre of the core (for a symmetric waveguide) to the point at which the field strength drops to 1/e. The functional form of the MFD is defined as:

\[ MFD = h + \frac{2}{\gamma} \cdot \]

(8.8)
Appendix C

The mode calculation carried out by LaserMod uses a Ritz iteration method to solve Maxwell’s equations in one dimension (the growth direction) across the layer interfaces. The Ritz method allows a direct approximate solution to differential equations constrained by boundary conditions. It is a form of finite element method and so deals with continuous functions with an infinite number of eigenstates by subdividing the problem into “finite elements” which can be evaluated individually and recombined to find a solution over the region of interest. The problem is solved by the minimisation of an associated error function, so in the case of determining the optical field of a guided mode the minimum of energy is found. An initial test wave function is chosen to meet the known boundary conditions detailed in Appendix D as well as any other relevant physical constraints (optical fields must have a non-negative magnitude for example) and the unknown parameters are treated as variables in order to minimise the energy of the system. Further details of the Ritz method can be found in reference [127].

Solving Maxwell’s equations along side a similar analysis to that detailed in section 6.1.1 gives the transverse electric field magnitude at each point on the grid across the structure, allowing the full transverse optical mode to be assembled. The confinement factor is then determined by summing the optical field intensity over all layers set as “active” (the QWs) and dividing by the sum of the optical field intensity over the full extent of the structure. The device structure is then saved as a text file that is called as an input when an operation such as “mode calculation” is executed from the user interface. Rsoft has a built in “parameter scan” function, which will vary a given single parameter e.g. the thickness of a certain layer over a range of values and with an interval set by the user. In the case of this investigation the P and N side cladding (BGaP) and barrier regions (BGaAsP) were varied over the ranges 0-5 \( \mu \)m in 25 nm steps and 0-1 \( \mu \)m in 50 nm steps respectively. The problem with this function is that it only allows one parameter to be varied at once. So in order to cover all possible combinations of layer thicknesses a MatLab script was written to control LaserMod. MatLab can be used to run any windows application by sending instructions directly to the command line. LaserMod has been designed in such a way as to allow its individual features such as parameter scan to be called individually from the command line. Thus it is possible to create a simple script to change a parameter in the structure text file and then run the parameter scan in a loop until all possibilities have been tried. Additional code written in MatLab collects the relevant information such as the confinement factor from the Rsoft output text files, determines which structure maximizes the confinement factor and finally plots the index profile, envelope function and confinement factor as a function of the variable changed by the script.

Appendix D

The post structure #16195 devices in addition to not lasing were found to have particularly poor IV characteristics. Figure 8.2 Shows an example IV (left) and ln(I) vs V (right) for a representative device of structure #25486, which is a reproduction of #16195. Also shown is a commercial grade InGaAs communications laser for comparison. The Turn on voltage is large, much greater than the bandgap energy, which it would be
expected to be approximately equal to. Turn on represents a transition from an exponential growth to an approximately linear phase (Ohmic) and so on a log scale a transition from linear to logarithmic is expected at turn on, as is the case for the InGaAs device. For the GaNAsP device however, there is no well defined turn on. It was thought that these IV characteristics might be related to un-optimised processing, leading to the SEM study of the contacts shown in section 6.2.

Figure 8.2: CW IV characteristics for both a commercial InGaAs laser and a representative GaNAsP device. Left: IV. Right: ln(I) vs V.

Appendix E

Figure 8.3 depicts the extrapolation from the approximately linear low injection limit of the LI curves to threshold to approximate the the radiative current. This method assumes that at low injection radiative recombination dominates and therefore \( I \approx I_{rad} \) and that the spontaneous emission rate is proportional to the radiative current, \( r_{sp} \propto I_{rad} = eVB\bar{n}^2 \). This extrapolation is therefore only possible in devices that are reasonably efficient so that there is some injection level for which the a linear region is found.
Figure 8.3: The extrapolation from the low injection, approximately linear regime where radiative recombination should dominate, based on the approximation \( r_{sp} \propto I_{rad} = eVBn^2 \). Extrapolating to threshold gives the approximate pure spontaneous emission intensity relative to the facet emission intensity.
References


