Towards Room-Temperature Terahertz Quantum Cascade Lasers: Directions and Design

by

Chun Wang Ivan Chan

B.ASc, University of British Columbia (2008)
S.M., Massachusetts Institute of Technology (2010)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2015

© Massachusetts Institute of Technology 2015. All rights reserved.

Author .................................................................
Department of Electrical Engineering and Computer Science
January 30, 2015

Certified by ..........................................................
Qing Hu
Professor of Electrical Engineering
Thesis Supervisor

Accepted by ..........................................................
Leslie A. Kolodziejski
Chairman, Department Committee on Graduate Theses
Towards Room-Temperature Terahertz Quantum Cascade Lasers: Directions and Design

by

Chun Wang Ivan Chan

Submitted to the Department of Electrical Engineering and Computer Science on January 30, 2015, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering

Abstract

Terahertz Quantum Cascade Lasers (THz QCLs) are arguably the most promising technology today for the compact, efficient generation of THz radiation. Their main limitation is that they require cryogenic cooling, which dominates their ownership cost. Therefore, achieving room-temperature operation is essential for the widespread adoption of THz QCLs. This thesis analyzes the limitations of THz QCL maximum lasing temperature ($T_{\text{max}}$) and proposes solutions.

THz QCL $T_{\text{max}}$ is hypothesized to be limited by a fundamental trade-off between gain oscillator strength $f_{ul}$ and upper-level lifetime $\tau$. This so-called “$f_{ul}\tau$ tradeoff” is shown to explain the failure of designs which target $\tau$ alone. A solution is proposed in the form of highly diagonal (low $f_{ul}$) active region design coupled with increased doping. Experimental results indicate the strategy to be promising, but heavily doped designs are shown to suffer band-bending effects which may deteriorate performance.

In order to treat these band-bending effects, which are typically neglected in previous THz QCL designs, a fast transport simulation tool is developed. Scattering integrals are simplified using the assumption of thermalized subbands. Results comparable to ensemble Monte Carlo are achieved at a fraction of the computational expense. Carrier leakages to continuum states are also investigated, although they are found to have little effect.

Other work in this thesis includes the optimization of double-metal THz waveguides to enable $T_{\text{max}} \sim 200$ K, a current world record. Furthermore, laser designs to investigate the leakages of carriers to high-energy subbands and continuum states were fabricated and tested; such parasitic leakages are suggested to be small. Finally, the design of gain media for applications is examined, notably the development of 4.7 THz gain media for OI line detection in astrophysics, and the development of broadband heterogeneous gain media for THz comb generation.

Thesis Supervisor: Qing Hu
Title: Professor of Electrical Engineering
Acknowledgments

During my time at MIT, I have had the honor of working with many outstanding mentors and colleagues both in and outside the Terahertz and Millimeter-wave Devices Group. I thank them here for their generosity.

I thank my supervisor Prof. Qing Hu for his guidance and support over the last 6 years. Of all his qualities as a researcher, his relentless drive is what stands out to me the most. While his optimism has often been at odds with my pragmatism (he unfairly calls it “pessimism”), it also inspired me to persevere through the equally relentless challenges and frustrations of research.

I am grateful to Dr. Tsung-Yu (Wilt) Kao and Dr. David Burghoff, my long-time labmates. They are two of the most brilliant men I’ve ever had the pleasure of working with, and I was forever bouncing ideas off of them. In particular, I thank Wilt for his mentorship during my early days in the cleanroom, and Dave’s genius in coming up with solutions to problems around the lab.

Dr. Amir Tavallaee was a pleasure to work with, always ready to share his numerous talents and breadth of knowledge. Whether it was a EM problem or a fabrication issue that I had, I knew I could reliably count on Amir to field a useful suggestion.

To Indrasen Bhattacharya, Shengxi Huang, and Ningren Han, thank you for teaching me how to teach, although I’m sorry you ended up being my guinea pigs in the process. I appreciate your tolerance of my shortcomings as a mentor and instructor.

My other great labmates include Dr. Asaf Albo, Yang Yang and Ali Khalatapour.

From my lab’s alumni, I thank Prof. Sushil Kumar for showing me the ropes of the QCL business when he was still a post-doc and I was a fledging graduate student. Dr. Qi Qin and Dr. Alan Lee were also very helpful during those days. Qi frequently helped me with equipment in the lab, and Alan stands out in my mind to this day as the archetype of the well-rounded individual.

Much of my work was in the cleanroom facilities of the Microsystems Technology Laboratory (MTL) at MIT. I thank MTL’s outstanding staff members for their maintenance of the equipment crucial to this work. In particular, I thank Dennis Ward,
Dave Terry, Bob Bicchieri, Eric Lim, and Donal Jamieson for going out of their way the time and time again to save me from belligerent and temperamental machinery.

A great deal of what I know about semiconductor fabrication comes from my tenure as a student member of the MTL Policy and Technology Committee. From this experience, I am particularly thankful to Dr. Vicky Diaduk and Dr. Jorg Scholvin for their consistently excellent advice, first on process, and later on career.

Other faculty at MIT: I thank Prof. Isaac Chuang for being my academic advisor at MIT. He gave me some much needed perspective on how my PhD education fit into the grand scheme of things (life and career). We met only twice a year, yet he made each time count; I always walked away with much to contemplate. I also thank Profs. Rajeev Ram and Dirk Englund for taking the time to serve on my PhD committee.

Outside of MIT, I am thankful to Dr. John Reno and his team at Sandia National Laboratories, who provided the abundant MBE growth crucial to this thesis.

I’ve also had many happy collaborations with researchers at the National Research Council of Canada and the University of Waterloo, including Dr. Emmanuel Dupont, Dr. Saeed Fathololoumi, Dr. Seyed Ghasem Razavipour, Prof. Dayan Ban, Prof. Zbig Wasilewski, and Prof. H. C. Liu. Prof. Liu unfortunately passed away part-way through my studies, in October 2013. May he rest in peace, secure in the tremendous scientific legacy that he leaves behind.

From the broader community of intersubband researchers, I thank Prof. Andreas Wacker and Prof. Masamichi Yamanishi for taking the time to discuss the finer points of intersubband transport with me during my occasional excursions to the ITQW/IQCLSW conferences. I very much enjoyed these conversations, and admire how freely these men share their knowledge.

To my family, for their consistent love and support through my life; my grandmother Mary, who taught me kindness; my mother Janice, who taught me responsibility; my sister Ida, who taught me independence; and father Nick, who taught me ambition: thank you for everything. I promise I will never stop trying to be a better person.

Research is expensive, and I am grateful to the funding provided by NASA and
NSF during my time at MIT. The United States is fortunate to have institutions such as these.

Finally, and most importantly, I thank Xiaowei, for bringing more joy to my life during my last two years at MIT than all the preceding years combined. I pray I will always be worthy of you.
# Contents

1 Introduction
   1.1 Background and Motivation ............................................. 21
   1.2 A brief review of THz sources .......................................... 22
   1.3 The Terahertz Quantum Cascade Laser ................................. 24
   1.4 Limitations of the THz QCL .............................................. 26
   1.5 Thesis objective and overview ......................................... 26

2 Transport theory for Quantum Cascade Lasers .......................... 29
   2.1 Computation of electronic states ....................................... 29
      2.1.1 Numerical solution .................................................. 31
   2.2 Tunneling ................................................................. 35
      2.2.1 The Kazarinov-Suris model ........................................ 35
      2.2.2 Application of KS tunneling to the 3-level direct phonon THz QCL .................................................. 37
      2.2.3 Second-order tunneling ............................................. 42
      2.2.4 The dephasing time .................................................. 45
   2.3 Single-body intersubband scattering .................................. 45
      2.3.1 Definitions of the rates ............................................. 46
      2.3.2 Transfer rates ....................................................... 48
      2.3.3 Cooling rate ......................................................... 50
      2.3.4 Heating rate ......................................................... 50
      2.3.5 Matrix elements for single body scattering ....................... 51
      2.3.6 Intersubband gain ................................................... 55
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4 Two-body intersubband scattering</td>
<td>56</td>
</tr>
<tr>
<td>2.4.1 Transfer rate</td>
<td>58</td>
</tr>
<tr>
<td>2.4.2 Cooling rate</td>
<td>62</td>
</tr>
<tr>
<td>2.4.3 Heating rate</td>
<td>63</td>
</tr>
<tr>
<td>2.5 Interactions between bound subbands and continuum states</td>
<td>67</td>
</tr>
<tr>
<td>2.5.1 Distinction between bound and continuum states</td>
<td>68</td>
</tr>
<tr>
<td>2.5.2 Mathematical description of continuum states</td>
<td>69</td>
</tr>
<tr>
<td>2.6 Single-body scattering between bound subbands and the continuum</td>
<td>70</td>
</tr>
<tr>
<td>2.6.1 Continuum-to-bound capture through LO phonon emission</td>
<td>71</td>
</tr>
<tr>
<td>2.6.2 Bound-to-continuum escape through LO phonon absorption</td>
<td>77</td>
</tr>
<tr>
<td>2.6.3 Verification of detailed balance</td>
<td>77</td>
</tr>
<tr>
<td>2.7 Two-body interactions between bound subbands and the continuum</td>
<td>78</td>
</tr>
<tr>
<td>2.7.1 Intersubband impact ionization</td>
<td>80</td>
</tr>
<tr>
<td>2.7.2 Intersubband Auger recombination</td>
<td>87</td>
</tr>
<tr>
<td>2.7.3 Verification of detailed balance</td>
<td>89</td>
</tr>
<tr>
<td>2.8 Conclusion</td>
<td>89</td>
</tr>
<tr>
<td>3 Transport modeling for Quantum Cascade Lasers</td>
<td>91</td>
</tr>
<tr>
<td>3.1 Formalisms for transport</td>
<td>92</td>
</tr>
<tr>
<td>3.2 Choice of model and physical assumptions</td>
<td>93</td>
</tr>
<tr>
<td>3.3 Computational basis</td>
<td>94</td>
</tr>
<tr>
<td>3.4 Computation of matrix elements for scattering</td>
<td>96</td>
</tr>
<tr>
<td>3.4.1 Coulombic form factors for electron-electron and LO phonon</td>
<td>97</td>
</tr>
<tr>
<td>scattering</td>
<td></td>
</tr>
<tr>
<td>3.4.2 Coulombic form factors for impurity scattering</td>
<td>100</td>
</tr>
<tr>
<td>3.4.3 Screening and the Coulombic matrix elements</td>
<td>100</td>
</tr>
<tr>
<td>3.5 Determination of populations</td>
<td>102</td>
</tr>
<tr>
<td>3.6 Determination of subband temperatures</td>
<td>102</td>
</tr>
<tr>
<td>3.6.1 Determination of the continuum temperature and mobility</td>
<td>103</td>
</tr>
<tr>
<td>3.7 Band bending</td>
<td>104</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Ti/Au versus Ta/Cu waveguides</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Effects of the ohmic contacting layer</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Waveguide optimization to achieve $T_{\text{max}} \sim 200$ K</td>
</tr>
<tr>
<td>5.4.4</td>
<td>Ti/Au vs Ta/Au waveguides</td>
</tr>
<tr>
<td>5.5</td>
<td>Investigations of alternative waveguide structures</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Experimental design</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Experimental results</td>
</tr>
<tr>
<td>5.6</td>
<td>Comparison to experiment</td>
</tr>
<tr>
<td>5.7</td>
<td>Conclusions</td>
</tr>
<tr>
<td>6</td>
<td>Effects of diagonality on optical gain</td>
</tr>
<tr>
<td>6.1</td>
<td>Trade-off between lifetime and oscillator strength ($f_{ul}\tau$)</td>
</tr>
<tr>
<td>6.2</td>
<td>Modeling of 3-well quantum cascade lasers</td>
</tr>
<tr>
<td>6.3</td>
<td>Overcoming $f_{ul}\tau$: the need for increased doping in THz QCLs</td>
</tr>
<tr>
<td>6.4</td>
<td>First studying of doping: NRC-V812 vs. NRC-V812-M1</td>
</tr>
<tr>
<td>6.5</td>
<td>Second study of doping: OWI210H-M series</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Poisson effects in highly doped THz QCLs</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Comparison to simulation</td>
</tr>
<tr>
<td>6.5.3</td>
<td>High doping and superdiagonality: OWI205H-M7</td>
</tr>
<tr>
<td>6.6</td>
<td>Other effects of doping</td>
</tr>
<tr>
<td>6.7</td>
<td>Elimination of Poisson effects using two-well designs</td>
</tr>
<tr>
<td>6.8</td>
<td>Conclusion</td>
</tr>
<tr>
<td>7</td>
<td>High Energy Parasitics and the Continuum</td>
</tr>
<tr>
<td>7.1</td>
<td>Elimination of high energy parasitics through ground state designs</td>
</tr>
<tr>
<td>7.1.1</td>
<td>Design of a 3-level ground-state design (OWIGS271)</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Experimental results of the 3-level ground-state design</td>
</tr>
<tr>
<td>7.1.3</td>
<td>Design of a 4-level ground-state design</td>
</tr>
<tr>
<td>7.1.4</td>
<td>Experimental results of the 4-level ground-state design</td>
</tr>
<tr>
<td>7.1.5</td>
<td>Comparison of OWIGS271 and TWIGS254, and other discussion</td>
</tr>
<tr>
<td>7.1.6</td>
<td>$f_{ul}\tau$ tradeoff in Ground State Designs</td>
</tr>
</tbody>
</table>
### 7.2 Suppression of leakages to continuum using tall barrier designs

#### 7.2.1 Design of a tall injector THz QCL (NRC-V775C)

- Page 192

#### 7.2.2 Experimental results of tall injector design

- Page 193

#### 7.2.3 Results for an all tall barrier design

- Page 197

#### 7.2.4 Conclusion

- Page 198

### 8 Terahertz QCL Gain Media for Practical Applications

#### 8.1 Terahertz QCLs for oxygen line detection

- Page 201
  - 8.1.1 4.7 THz QCLs based on 3-well resonant phonon active regions
  - Page 202
  - 8.1.2 4.7 THz QCLs based on 4-well resonant phonon designs
  - Page 211
  - 8.1.3 Conclusion
  - Page 213

#### 8.2 Broadband heterogeneous terahertz QCLs

- Page 216
  - 8.2.1 Interlaced heterogeneous terahertz QCL
  - Page 216
  - 8.2.2 Charge imbalance in heterogeneous structures
  - Page 218
  - 8.2.3 Current matching in heterogeneous structures
  - Page 220
  - 8.2.4 Segregated Hetergeneous Gain Media
  - Page 220

#### 8.3 Conclusions

- Page 221

### 9 Summary

#### 9.1 Contributions of this thesis

- Page 227

#### 9.2 Directions for future work

- Page 228

### A $k \cdot p$ parameters

- Page 243

### B Sample growth sheet

- Page 247

### C Additional processing information

- Page 249
# List of Figures

1-1 The “terahertz gap” in the electromagnetic spectrum . . . . . . . . . 21
1-2 Power and frequency of terahertz sources . . . . . . . . . . . . . . . . 23
1-3 Intersubband vs. interband transitions . . . . . . . . . . . . . . . . 25
1-4 Schematic of QCL operation. . . . . . . . . . . . . . . . . . . . . . . 25
1-5 Evolution of THz QCL maximum lasing temperature over time. . . . . 27
1-6 Maximum lasing temperature versus lasing frequency. . . . . . . . . . 27

2-1 Matrix structure in spectral-element based $k \cdot p$ method . . . . . . 34
2-2 Typical values of the Kazarinov-Suris tunneling time $\tau^*$ at resonance. 38
2-3 Schematic model of a three level THz QCL . . . . . . . . . . . . . . . . 39
2-4 Schematic illustration of second-order tunneling. . . . . . . . . . . . 43
2-5 Single-body scattering mechanisms. . . . . . . . . . . . . . . . . . . . 46
2-6 Changes in subband statistics upon scattering . . . . . . . . . . . . . 47
2-7 Pair-wise electron electron scattering between two subbands. . . . . . 57
2-8 The local barrier height. . . . . . . . . . . . . . . . . . . . . . . . . . . 70
2-9 LO phonon interactions with the continuum. . . . . . . . . . . . . . . 71
2-10 Intersubband impact ionization and Auger recombination. . . . . . . 79

3-1 Schematic of the matrix structure of the Hamiltonian . . . . . . . . . 95
3-2 Sample flatband basis states for transport simulation. . . . . . . . . . 96
3-3 Flowchart detailing major steps of QCL simulation. . . . . . . . . . . 107
3-4 Band diagram of OWI222G. . . . . . . . . . . . . . . . . . . . . . . . 108
3-5 Computed vs experiment IVs of OWI222G. . . . . . . . . . . . . . . . . 109
3-6 Spurious absorption in FL183S. . . . . . . . . . . . . . . . . . . . . . . 111
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-7</td>
<td>Computed vs. experimental IV curves of FL183S</td>
</tr>
<tr>
<td>3-8</td>
<td>Subband temperatures in $T_{\text{max}}$ record-holding THz QCL design.</td>
</tr>
<tr>
<td>3-9</td>
<td>Scattering velocity in $T_{\text{max}}$ record-holding THz QCL design.</td>
</tr>
<tr>
<td>4-1</td>
<td>Schematic depiction of epitaxial growth structure.</td>
</tr>
<tr>
<td>4-2</td>
<td>Major steps in double-metal waveguide formation for Terahertz Quantum Cascade laser fabrication. Steps read top-to-bottom, left-to-right.</td>
</tr>
<tr>
<td>4-3</td>
<td>Optical and SEM photographs of finished devices.</td>
</tr>
<tr>
<td>4-4</td>
<td>A mounted QCL die photographed next to an American penny.</td>
</tr>
<tr>
<td>5-1</td>
<td>Gain versus temperature for a particular QCL.</td>
</tr>
<tr>
<td>5-2</td>
<td>Double metal waveguide geometry</td>
</tr>
<tr>
<td>5-3</td>
<td>Waveguide loss vs. frequency for different waveguide metallizations</td>
</tr>
<tr>
<td>5-4</td>
<td>Effect of the removing the top contact layer on double-metal waveguide loss.</td>
</tr>
<tr>
<td>5-5</td>
<td>Experimental data for $T_{\text{max}}$ record holding THz QCL device, reproduced from ref. [1].</td>
</tr>
<tr>
<td>5-6</td>
<td>Comparison of 10 nm Ta/Au and 5 nm Ti/Au waveguides.</td>
</tr>
<tr>
<td>5-7</td>
<td>Losses in alternative waveguide structures.</td>
</tr>
<tr>
<td>5-8</td>
<td>Experimental comparison of alternative waveguide structures</td>
</tr>
<tr>
<td>6-1</td>
<td>Thermally activated LO phonon scattering</td>
</tr>
<tr>
<td>6-2</td>
<td>Schematic of analytical model for QCL transport for a 3-well, 4-level QCL</td>
</tr>
<tr>
<td>6-3</td>
<td>Maximum lasing temperature versus oscillator strength</td>
</tr>
<tr>
<td>6-4</td>
<td>The effect of increasing doping on the $f\tau$ trade-off.</td>
</tr>
<tr>
<td>6-5</td>
<td>Band diagram of design NRC-V812 and NRC-V812-M1</td>
</tr>
<tr>
<td>6-6</td>
<td>Experimental data for NRC-V812 (VB0523).</td>
</tr>
<tr>
<td>6-7</td>
<td>Experimental data for NRC-V812-M1 (VB0523).</td>
</tr>
<tr>
<td>6-8</td>
<td>Band diagram of OWI210H (VB0373).</td>
</tr>
<tr>
<td>6-9</td>
<td>SIMS depth profiling of dopants in OWI210H-M3 (VB0605</td>
</tr>
<tr>
<td>Page</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>6-10</td>
<td>Experimental data for OWI210H-M3 (VB0605).</td>
</tr>
<tr>
<td>6-11</td>
<td>Experimental data for OWI210H-M4 (VB0610).</td>
</tr>
<tr>
<td>6-12</td>
<td>Experimental data for OWI210H-M5 (VB0612).</td>
</tr>
<tr>
<td>6-13</td>
<td>Experimental data for OWI210H-M6 (VB0609).</td>
</tr>
<tr>
<td>6-14</td>
<td>Comparison of data from OWI210H-M devices.</td>
</tr>
<tr>
<td>6-15</td>
<td>Bandstructure of OWI210H-M series devices at injection resonance.</td>
</tr>
<tr>
<td>6-16</td>
<td>Computed IVs and LI's at for OWI210H-M series of devices.</td>
</tr>
<tr>
<td>6-17</td>
<td>Band diagram of OWI205H-M7.</td>
</tr>
<tr>
<td>6-18</td>
<td>Experimental data for OWI205H-M7 (VB0611).</td>
</tr>
<tr>
<td>6-19</td>
<td>Schematic band-diagram of a two-well THz QCL.</td>
</tr>
<tr>
<td>6-20</td>
<td>$f_d\tau$ product vs. oscillator strength for a two-well THz QCL.</td>
</tr>
<tr>
<td>7-1</td>
<td>One module band diagram for OWIGS271.</td>
</tr>
<tr>
<td>7-2</td>
<td>Experimental data for OWIGS271.</td>
</tr>
<tr>
<td>7-3</td>
<td>One module band diagram for TWIGS254.</td>
</tr>
<tr>
<td>7-4</td>
<td>Experimental data for TWIGS254 clad.</td>
</tr>
<tr>
<td>7-5</td>
<td>$f_d\tau$ product as a function of transition height.</td>
</tr>
<tr>
<td>7-6</td>
<td>Band diagrams for NRC-V775C and NRC-V775A</td>
</tr>
<tr>
<td>7-7</td>
<td>Experimental LIVT and spectral data for NRC-V775C and NRC-V775A.</td>
</tr>
<tr>
<td>7-8</td>
<td>Comparison of threshold current rise between NRC-V775C and NRC-V775A.</td>
</tr>
<tr>
<td>7-9</td>
<td>OWI230T band diagram and experimental data.</td>
</tr>
<tr>
<td>8-1</td>
<td>Band diagram of OWI223R.</td>
</tr>
<tr>
<td>8-2</td>
<td>Experimental data for OWI223 (VB0493).</td>
</tr>
<tr>
<td>8-3</td>
<td>Experimental data for OWI223 (VB0492).</td>
</tr>
<tr>
<td>8-4</td>
<td>Experimental data for OWI223 (VB0488).</td>
</tr>
<tr>
<td>8-5</td>
<td>Band diagram of OWI223R.</td>
</tr>
<tr>
<td>8-6</td>
<td>Experimental data for OWI219R-M1 (VB0524).</td>
</tr>
<tr>
<td>8-7</td>
<td>Experimental data for OWI219R-M1 (VB0540).</td>
</tr>
<tr>
<td>8-8</td>
<td>Experimental data for OWI219R-M1 (VB0541).</td>
</tr>
</tbody>
</table>
8-9  Band diagram of FL182R-M6.                            212
8-10 Band diagram of FL181R-M7.                            212
8-11 Experimental data for FL182R-M6 (VB0520).              214
8-12 Experimental data for FL181R-M7.                      215
8-13 Band diagrams of constituent designs of heterogenous gain medium FLR4.                                217
8-14 Gain curve of FLR4 based on sum-of-Lorentzians estimate, assuming equal gain from each gain medium.       218
8-15 Experimental data for heterogenous gain medium FLR4 (VB0621).                                       219
8-16 Band diagrams for constituent designs of OWIE3-M1.     222
8-17 Computational IV results for OWIE3-M1                    223
8-18 Experimental data for heterogenous gain medium OWIE3-M1 (VB0700)                                    224
A-1  Al$_x$Ga$_{1-x}$As/GaAs conduction band offset          245
# List of Tables

5.1 Resistivities at 20°C for select waveguide metals, as well as Drude parameters. ....................................................... 140
5.2 Summary of experimental results ........................................... 147

6.1 Summary of parameters used in the semi-analytical model of a 4-level THz described in figure 6-2 ........................................... 155
6.2 Designs represented in figure 6-20. These are generated to have $E_{32} = 15$ meV, $E_{21} = 36$ meV, and $\Delta_{1'3} = 2.5$ meV. Layer thicknesses are given in nm. Band-bending is neglected. .............................. 180

A.1 $k \cdot p$ parameters used for Al$_x$Ga$_{1-x}$As ................................. 244
Chapter 1

Introduction

1.1 Background and Motivation

The electromagnetic spectrum as known today is depicted in Figure 1-1. From DC up to about 1 THz is the domain of electronic technology (radio, wifi, cellular and so forth), and from the X-rays down to about 10 THz is the domain of photonics (lasers and LEDs, primarily). The 1 to 10 THz region is the so-called “THz Gap,” a technological dead-zone where excellent detectors exist, but also a dearth of good radiation sources. At the time of writing of this thesis, there is an explosion of interest in exploiting this terahertz gap. Merely three years ago, in 2011, the Institute of Electrical and Electronics Engineers (IEEE) inaugurated the IEEE Transactions on Terahertz Science and Technology for the study of this field. Terahertz radiation has many useful prop-

Figure 1-1: The “terahertz gap” in the electromagnetic spectrum. Few natural sources of radiation exist in this range.
erties that make it attractive for certain applications, ranging from chemical sensing and spectroscopy to security and astronomical imaging (see, for example, [2–7]).

A detailed examination of the merits and shortcomings of various proposals for THz technology is beyond the scope of this work, but they are united in their need for powerful sources of THz radiation. The lack of good sources is the single greatest obstacle to THz science. The broad aim of this thesis is to supply such sources to advance the development of THz science and technology.

1.2 A brief review of THz sources

The ideal radiation source would be compact, powerful, efficient and capable of broad spectral coverage. No technology exists today that attains all of these criteria simultaneously.

An overview of present day THz sources is presented in figure 1-2. As categorized by Armstrong, [8] there are roughly four groups of THz sources, these being vacuum electronics, solid-state electronics, laser sources, and non-laser photonic sources. Vacuum electronics and solid-state electronics are dominant in the sub-THz and low-THz ranges. Above 1 THz or so, however, their power drops off rapidly. The mid- and high-THz range is therefore dominated by photonics.

In this regard, optically pumped molecular gas lasers have excellent power, and a wide range of available frequencies. Unfortunately, their spectral coverage is limited to discrete lines corresponding to certain molecular vibration modes. Gas lasers are also rather large table top devices, with a typical laser cavity being around a 1-3 m long.

For compact sources (mm to cm scale), options are then limited to the terahertz quantum cascade laser, and optical down-conversion based on optical rectification, difference frequency generation (DFG), or photomixing (beat frequency generation). Both categories of devices are spectrally agile, capable of continuous coverage in the ~1-5 THz range. However, QCLs have the clear advantage in power and efficiency. For example, the average power of a well-designed QCL gain medium clad in a 3rd
Figure 1-2: Representative power and frequencies for existing terahertz source technologies, reproduced from ref. [8]. Quantum Cascade Lasers and molecular gas lasers standard from all competing technologies in the high-THz spectrum.
order distribute feedback waveguide can output average powers in the 10 mW range with narrow beam-pattern with $\sim 0.5\%$ wallplug efficiency. [9] Pulsed QCL sources yielding $> 1$ W of THz power have also been recently demonstrated. [10] In comparison, most down-conversion based sources output only $\mu$W-level powers. Thus, a case can be made that THz QCLs remain the most promising source of THz radiation today, at least on the high-frequency side of the THz gap.

The interested reader is referred to several excellent reviews of existing terahertz sources for more information. [8,11]

### 1.3 The Terahertz Quantum Cascade Laser

The quantum cascade laser (QCL) is a semiconductor laser that generates optical amplification through intersubband transitions. This contrasts conventional semiconductor lasers, which operate based on interband transitions. Since naturally occurring bandgaps tend to be above $\sim 60$ meV (15 THz; e.g. in lead-salt lasers [12]), finding the right semiconductor material for building an interband THz laser is extremely challenging. The advantage of intersubband transitions is that their energetic spacing can be engineered through adjusting quantum well widths and barrier heights (figure 1-3). A THz QCL is formed of hundreds of such wells, arrange into a superlattice. The material system is typically AlGaAs/GaAs, although other materials systems have been explored (e.g. InGaAs/InAlAs [13] and InGaAs/GaAsSb [14]).

In conventional quantum well lasers, population inversion is achieved through supplying holes in the valence band using a $p$-type material. There is no equivalent to a $p$-type material in a quantum cascade laser, as the all transport happens in the conduction band. Instead, population inversion is achieved through recycling the carriers from one superlattice period into the next, usually through resonant tunneling injection of carriers into the next upper laser level. The overall process is depicted in figure 1-4.

Kazarinov and Suris first proposed the QCL concept in 1971 [15]. The midinfrared (MIR) QCL was first demonstrate in 1994 by Faist et al., [16] and the THz QCL in...
Figure 1-3: Intersubband versus interband transitions. Unlike interband transition energies ($\hbar\omega_2$), which are essentially restricted to the bandgap of the well, intersubband transition energies ($\hbar\omega_1$) can be engineered through adjusting the well width and barrier heights.

Figure 1-4: Schematic of QCL operation. In principle, each electron emits one photon in each superlattice period before moving into the next period.
1.4 Limitations of the THz QCL

Unfortunately, THz QCLs presently lase only at cryogenic temperatures (< 200 K). Cryogenic cooling equipment is typically in the 10’s of cm to meter-scale in size, and extremely power hungry. This compromises the compactness and power efficiency advantages. For the widespread adoption of THz QCL technology to occur, this particular technological barrier must be removed.

Since their inception, MIR QCLs have advanced in leaps and bounds, to the extent that they are the gold-standard for MIR generation. Powerful and efficient sources of MIR QCLs operate at room temperature, and are sold as commercial products (representative vendors as of 2014 include Daylight Solutions in the United States, Hamamatsu Photonics in Japan, and Alpes Lasers in Switzerland).

In contrast, THz QCLs advanced rapidly up until 2005, but progress since then has stagnated. Figure 1-5 shows the evolution of maximum lasing temperature ($T_{\text{max}}$) since their first demonstration in 2002. A crude extrapolation based on the current rate of improvement suggests that a room temperature laser may not be seen until a decade and a half from now. Figure 1-6 also shows a map of published THz QCL data on a plot of $T_{\text{max}}$ vs frequency.

1.5 Thesis objective and overview

This thesis aimed to break this stagnation, and enable the room temperature operation of THz QCLs. Ultimately, my efforts failed, but my hope is that the results herein can guide future research towards the goal of room temperature operation.

The remainder of this thesis is organized as follows.

- Chapter 2 describes the physics of electron transport in THz QCLs.
- Chapter 3 draws upon the theoretical foundation of chapter 2 to build a quantitative modeling tool for QCL design.
Figure 1-5: Evolution of THz QCL maximum lasing temperature over time. In order from earliest to latest, the points are taken from [17], [19], [20], [21], [22], [23], [24], [1].

Figure 1-6: A map of available QCL frequencies and their respective maximum lasing temperatures. This is figure 1b from [25], updated with data on lasers published since 2007. A reasonable attempt to be comprehensive has been made, but due to the vastness of the QCL literature, I apologize for any errors of omission. Data is organized according to mode of operation (pulsed or continuous wave (CW)) and design type. Guide to designs: CSL - chirped superlattice; BTC - bound-to-continuum; RP - resonant-phonon; DP - direct-phonon; IDP - indirect pump (also called “scattering assisted”).
• Chapter 4 describes the experimental methods used in THz QCL fabrication and characterization.

• Chapter 5 describes efforts to reduce losses in THz QCL waveguides.

• Chapter 6 describes the hypothesis that THz QCL temperature performance is limited by a fundamental trade-off between laser oscillator strength ($f_{ul}$) and upper laser level lifetime ($\tau$). This $f_{ul}\tau$ hypothesis is the central result of this thesis. Increased doping is presented as a possible suggestion to break this trade-off.

• Chapter 7 describes work performed on investigating other posited causes of lasing degradation versus temperature, namely high energy parasitics states.

• Chapter 8 changes track to describe work done on building THz gain media for specific applications (as opposed to the sole pursuit of higher operating temperatures).
Chapter 2

Transport theory for Quantum Cascade Lasers

The chapter assembles a toolkit for the theoretical analysis of terahertz quantum cascade laser (THz QCLs) QCLs. THz QCLs engineer gain between subband states in a semiconductor superlattice through a combination of coherent and noncoherent physics for their operation. In terms of electrical transport, these physics correspond to tunneling and scattering between subband states. Basic mathematical descriptions of these phenomena are presented.

2.1 Computation of electronic states

The most fundamental tool of THz QCL design and analysis is a subband bandstructure calculator. In this section, the $k \cdot p$ method is employed to calculate the electronic states in the semiconductor superlattice in the envelope function description. In keeping with standard practice, subband wavefunctions and energies are calculated at the subband edges ($k_x, k_y = 0$), and in-plane dispersion is modelled through a simple effective mass description.

Let $z$ be the superlattice growth direction. The electronics state can be described by the 8-band $k \cdot p$ Kane Hamiltonian (which includes the conduction, light-hole, heavy-hole, and splitoff bands and their spin interactions). In GaAs, the bands
are two-fold spin degenerate, leaving us with four bands. Furthermore, the matrix elements coupling the heavy-hole band to the other three bands vanish when both $k_x$ and $k_y$ are zero, so the heavy-hole band decouples. Therefore, we start with the $3 \times 3$ Hamiltonian

$$H = \begin{bmatrix}
E_c + p_z \frac{F}{2m_0} p_z & \sqrt{\frac{2}{3}} ip_z p_z & -\sqrt{\frac{1}{3}} ip_z p_z \\
-\sqrt{\frac{2}{3}} ip_z P & E_v - p_z \frac{(\gamma_1 + 2\gamma_2)}{2m_0} p_z & \sqrt{2} p_z \frac{\gamma_2}{m_0} p_z \\
\frac{1}{3} ip_z P & \sqrt{2} p_z \frac{\gamma_2}{m_0} p_z & E_{so} - p_z \frac{\gamma_1}{2m_0} p_z
\end{bmatrix} \quad (2.1)$$

The symbols are defined as follows.

- $E_c$ is the conduction band edge.
- $E_v$ is the valence band edge.
- $E_{so}$ is the split-off band edge.
- $F = 1 + 2F_K$ is a quantity related to the the Kane parameter $F_K$.
- $P$ is the Kane momentum matrix element.
- $\gamma_1$ and $\gamma_2$ are the modified Luttinger parameters.
- $m_0$ is the free electron mass.

Except for the electron mass, all of the above are functions of position $z$. Parameter values given in appendix [A].

Following the prescription of the envelope function method, we replace $p_z$ by the operator $-i\hbar \partial / \partial z$. Therefore, the energy spectrum is specified by the eigenvalue problem

$$\begin{bmatrix}
E_c - \hbar^2 \frac{\partial}{\partial z} \frac{F}{2m_0} \frac{\partial}{\partial z} \\
-\hbar \hbar \frac{\partial}{\partial z} P & E_v + \hbar^2 \frac{\partial}{\partial z} \frac{(\gamma_1 + 2\gamma_2)}{2m_0} \frac{\partial}{\partial z} & -\hbar^2 \sqrt{2} \frac{\partial}{\partial z} \frac{\gamma_2}{m_0} \frac{\partial}{\partial z} \\
\hbar \frac{1}{3} \frac{\partial}{\partial z} P & -\hbar^2 \sqrt{2} \frac{\partial}{\partial z} \frac{\gamma_2}{m_0} \frac{\partial}{\partial z} & E_{so} + \hbar^2 \frac{\partial}{\partial z} \frac{\gamma_1}{2m_0} \frac{\partial}{\partial z}
\end{bmatrix} \begin{bmatrix}
\phi_{cb} \\
\phi_{lh} \\
\phi_{so}
\end{bmatrix} = \begin{bmatrix}
\phi_{cb} \\
\phi_{lh} \\
\phi_{so}
\end{bmatrix} \quad (2.2)$$

30
Where $\phi_{cb, lh, so}$ are the envelope functions for the conduction, light-hole, and split-off bands, respectively. Equation 2.2 describes three coupled differential equations in 1D that can be solved numerically.

### 2.1.1 Numerical solution

Equation 2.2 can be solved using the spectral element method (SEM), which is essentially a variant of the finite element method. The details of this method are explained in refs. [29, 30]. Briefly, the method assumes that the computation domain can be divided into discrete elements, similar to the finite element method (FEM). Inside each element, however, SEM assumes an expansion into orthogonal polynomials that is typically much higher than typical FEM implementations. The SEM may be regarded as a very high polynomial order FEM. SEM can achieve better than an order of magnitude improvement in speed compared to the finite difference method.

Each QCL layer is taken to be one element. In the $i$-th QCL layer, we assume a Gauss-Legendre-Lobatto (GLL) pseudospectral expansion for the envelope functions. With $N_i$ collocation points in the $i$-th layer, we have

$$\phi_b \approx \sum_{n=1}^{N_i} \phi_{bn} f_n(z), \ b = cb, lh, so$$  \hspace{1cm} (2.3)

where the expansion coefficients $\phi_{bn}$ are equal to the values of $\phi_b$ are the $n$-th collocation point (that is, $\phi_{bn} = \phi_b(z_n)$). In the layer, the basis functions obey the relationships

$$\int_i dz f_n f_m = \delta_{nm}$$  \hspace{1cm} (2.4)
$$f_n(z_m) = \delta_{nm}$$  \hspace{1cm} (2.5)

where $\delta$ is the Kronecker delta function, and continuous space integration can be
approximated by the discrete sum

\[
\int_i dz h(z) \approx \sum_{p=1}^{N_i} w_p h(z_p) = \sum_{p=1}^{N_i} w_p h_p
\]  

(2.6)

where \(w_m\) are the Gaussian quadrature weights corresponding to the Legendre collocation points. \[29\]

The procedure goes as follows:

1. The expansions 2.3 are inserted into equations 2.2

2. Galerkin’s method is applied (integration on both sides against \(f_m\)). Derivatives of abruptly changing functions (such as the momentum matrix elements and Luttinger parameters) are removed using integration by parts, so that only derivatives of smooth functions remain.

3. The integrals of Galerkin’s method are approximated by equation 2.6

After much algebra, the system of coupled differential equations 2.2 can be written in block matrix form

\[
\begin{bmatrix}
A_{b-cb} & A_{b-lh} & A_{b-so} \\
A_{l-cb} & A_{l-lh} & A_{l-so} \\
A_{s-cb} & A_{s-lh} & A_{s-so}
\end{bmatrix}
\begin{bmatrix}
\phi_{b} \\
\phi_{lh} \\
\phi_{so}
\end{bmatrix}
= \begin{bmatrix}
W & 0 & 0 \\
0 & W & 0 \\
0 & 0 & W
\end{bmatrix}
\begin{bmatrix}
\phi_{b} \\
\phi_{lh} \\
\phi_{so}
\end{bmatrix}
+ \begin{bmatrix}
b_{cb} \\
b_{lh} \\
b_{so}
\end{bmatrix}
\]  

(2.7)

where the \(N_i \times 1\) vectors \(\phi_b = [\phi_{b1} \ldots \phi_{bN_i}]^T\) (\(b = cb, lh, so\)) have been defined and the \(N_i \times N_i\) matrix \(W\) is a diagonal matrix whose diagonal elements are given by the
Gaussian quadrature weights. The $N_i \times N_i$ matrices $A$ are given element-wise by

$$
\begin{align*}
\left[ A_{b-cb} \right]_{mn} &= E_{c,m} w_m \delta_{mn} + \frac{\hbar^2}{2m_0} \sum_{p=1}^{N_i} (D_{pm} w_p F_p D_{pn}) \\
\left[ A_{b-lh} \right]_{mn} &= \hbar \sqrt{\frac{2}{3}} w_m P_m D_{mn} \phi_n^{lh} = \left[ A_{h-lh} \right]_{nm} \\
\left[ A_{b-so} \right]_{mn} &= -\hbar \sqrt{\frac{1}{3}} w_m P_m D_{mn} \phi_n^{so} = \left[ A_{so-so} \right]_{nm} \\
\left[ A_{h-lh} \right]_{mn} &= E_{v,m} w_m \delta_{mn} - \frac{\hbar^2}{2m_0} \sum_{p=1}^{N_i} \left( D_{pm} w_p (\gamma_1 + 2\gamma_2) P_{p} D_{pn} \right) \\
\left[ A_{h-so} \right]_{mn} &= \sqrt{\frac{2}{3}} \hbar^2 m_0 \sum_{p=1}^{N_i} \left( D_{pm} w_p \gamma_2 P_{p} D_{pn} \right) = \left[ A_{so-lh} \right]_{nm} \\
\left[ A_{so-so} \right]_{mn} &= E_{so,m} w_m \delta_{mn} - \frac{\hbar^2}{2m_0} \sum_{p=1}^{N_i} \left( D_{pm} w_p \gamma_1 P_{p} D_{pn} \right)
\end{align*}
$$

where $E_{b,m} = E_b(z_m)$, $P_m = P(z_m)$, $F_p = F(z_p)$, $\gamma_p = \gamma(z_p)$ and the derivative matrix is $D_{mn} = \left[ \frac{\partial f_m (z)}{\partial z} \right]_{z_m}$. The vectors $\overline{b}$ define the boundary terms that arise from the integration by parts procedure. They are given element-wise by

$$
\begin{align*}
\left[ b_{cb} \right]_m &= \hbar^2 \left[ f_m \frac{F}{2m_0} \frac{\partial \phi_{cb}}{\partial z} \right]_{z_1}^{z_{N_i}} \\
\left[ b_{lh} \right]_m &= \hbar \sqrt{\frac{2}{3}} \sum_{n} \left( [f_m P \phi_{cb}]_{z_1}^{z_{N_i}} \right) - \hbar^2 \left[ f_m \frac{(\gamma_1 + 2\gamma_2) \partial \phi_{lh}}{2m_0 \partial z} \right]_{z_1}^{z_{N_i}} + \hbar^2 \sqrt{2} \left[ f_m \frac{\gamma_2 \partial \phi_{so}}{m_0 \partial z} \right]_{z_1}^{z_{N_i}} \\
\left[ b_{so} \right]_m &= -\hbar \sqrt{\frac{1}{3}} \left[ f_m P \phi_{cb} \right]_{z_1}^{z_{N_i}} + \hbar^2 \sqrt{2} \left[ f_m \frac{\gamma_2 \partial \phi_{lh}}{m_0 \partial z} \right]_{z_1}^{z_{N_i}} - \hbar^2 \left[ f_m \frac{\gamma_1 \partial \phi_{so}}{2m_0 \partial z} \right]_{z_1}^{z_{N_i}}
\end{align*}
$$

Note that $f_m (z_1) = \delta_{m1}$ and $f_m (z_{N_i}) = \delta_{mN_i}$, so the $[b]_m = 0$ except at the endpoints of the element.

To help understand the block matrices $A$, the matrix structure of $A_{b-cb}$ is schematically illustrated in figure 2-1 in connection to the analysis of a two-well QCL.

The boundary terms have been stated here only for completeness. Per usual procedure in finite-element methods, the internal boundary terms (boundary terms between two elements) cancel out when combining elements. This work uses Dirichlet
Figure 2-1: Matrix structure of the $A_{cb-cb}$ block of equation 2.7 for a two well quantum cascade laser. Shaded blocks inside $A_{cb-cb}$ denote non-zero matrix elements, with red for the barriers and blue for the wells. Adjacent blocks overlap (add together) at the corners.
boundary conditions \( \phi_{cb,lh,so} = 0 \) at simulation boundaries), so the external boundary terms may also be ignored. With all boundary terms eliminated, equation (2.7) forms an generalized eigenvalue problem, which can be solved through standard linear algebra libraries. Typically, eigenstates are computed over 1-3 periods of a QCL structure.

### 2.2 Tunneling

In THz QCLs, typically the barriers inside a module are thin enough that all wells are strongly coupled. Conversely, QCLs modules are typically separated by a thick quantum barrier (the injector barrier). Therefore, transport between QCL modules is better described through tunneling between weakly coupled wells. In the past, neglect of coherent effects between QCL modules has led to unphysical effects such as the current density of a QCL being independent of the injector barrier thickness. \([31]\)

#### 2.2.1 The Kazarinov-Suris model

This model was first introduced by Kazarinov and Suris in ref. \([15]\). Considered a localized basis of two subband coupled across a thick quantum barrier. Although technically a subband is comprised of many quantum states (corresponding to different in-plane momenta), each subband in a superlattice may be considered as an effective “0-dimensional” quantum level. The dynamics of this system in the density matrix formalism are described by Liouville’s equation,

\[
 i\hbar \frac{\partial}{\partial t} \rho = [H, \rho] + i\hbar \Gamma(\rho) \tag{2.17}
\]

where \( \rho \) is the \( 2 \times 2 \) density matrix, and the Hamiltonian is given by

\[
 H = \begin{bmatrix} E_1 & -V \\ -V & E_2 \end{bmatrix} \tag{2.18}
\]

In \( H \), \( E_1 \) and \( E_2 \) are the energies of the two subband edges, and \( V \) is the interaction between the two subbands induced by the coupling across the barrier. \( \Gamma(\rho) \) is the
superoperator representing scattering in the system, taken to possess a relaxation time form.

\[
\Gamma (\rho) = \begin{bmatrix}
\frac{-\rho_{11}}{\tau_1} + (\cdots) & -\frac{\rho_{12}}{\tau_1} \\
-\frac{\rho_{21}}{\tau_1} & -\frac{\rho_{22}}{\tau_2} + (\cdots)
\end{bmatrix}
\] (2.19)

In \( \Gamma (\rho) \), \( \tau_{1,2} \) represents the subband lifetimes, and \((\cdots)\) represents the gain or loss of electrons from other subbands (unimportant for this analysis). Expanding Liouville’s equation explicitly for the populations yields

\[
\frac{\partial}{\partial t} \rho_{11} = i\Omega (\rho_{21} - \rho_{12}) - \frac{\rho_{11}}{\tau_1} + (\cdots)
\] (2.20)

\[
\frac{\partial}{\partial t} \rho_{22} = i\Omega (\rho_{12} - \rho_{21}) - \frac{\rho_{22}}{\tau_1} + (\cdots)
\] (2.21)

(2.22)

where the abbreviations \( \Omega = V/\hbar \) and \( \omega_{ij} = (E_i - E_j)/\hbar \) have been used. In the equations for the populations, the terms \( \pm i\Omega (\rho_{21} - \rho_{12}) \) can be interpreted as population transfer due to the density matrix coherences; in other words, these terms correspond to the intersubband tunneling process.

The coherences are needed in order to calculate the tunneling rate. Instead of explicitly solving for the coherences, however, the key essence of the Kazarinov-Suris method is to approximate the density matrix coherence (off-diagonal terms) using the populations (on diagonal terms). [32] Returning to the Liouville equation, in the relaxation time approximation the coherences are described by

\[
\frac{\partial}{\partial t} \rho_{12} = -i\omega_{12}\rho_{12} - i\Omega (\rho_{11} - \rho_{22}) - \frac{\rho_{12}}{\tau_{||}}
\] (2.23)

\[
\frac{\partial}{\partial t} \rho_{21} = -i\omega_{21}\rho_{21} - i\Omega (\rho_{22} - \rho_{11}) - \frac{\rho_{21}}{\tau_{||}}
\] (2.24)

Assuming that a steady-state does exist, the steady-state coherences are thus given
by

\[ \rho_{12} = \frac{i\Omega (\rho_{11} - \rho_{22})}{i\omega_{12} + 1/\tau_{||}} \]  \hspace{1cm} (2.25)

\[ \rho_{21} = \frac{i\Omega (\rho_{22} - \rho_{11})}{i\omega_{21} + 1/\tau_{||}} \]  \hspace{1cm} (2.26)

and therefore, the tunneling rate is given by

\[ i\Omega (\rho_{12} - \rho_{21}) = -\frac{2\Omega^2 \tau_{||}}{1 + \omega^2_{12} \tau_{||}} (\rho_{22} - \rho_{11}) \]

\[ = -\frac{\rho_{22} - \rho_{11}}{\tau^*} \]  \hspace{1cm} (2.27)

The quantity \( \tau^* \) is the Kazarinov-Suris tunneling time (KS time), and represents the tunneling rate of electrons between subbands. The results generalize easily to multisubband system, so long as tunneling occurs primarily between two localized subbands (generalization to multisubband tunneling is straightforward but analytically cumbersome). The KS method dramatically simplifies transport calculations for intersubband tunneling, as it reduces quantum transport to rate equations. Typical values of \( \tau^* \) are plotted in figure 2-2.

### 2.2.2 Application of KS tunneling to the 3-level direct phonon THz QCL

In ref. [33] and [34], Kumar derives analytical formulas for QCL transport by explicitly solving the full density matrix equations. As an application of the section 2.2.1, this section shows that Kumar’s equations can be derived more simply using an effective rate equations approach.

**Below threshold**

In ref. [33], Kumar studies the 3-level QCL illustrated in figure 2-3 (at low temperatures, backfilling effects are ignored). The effective rates equations describing the
Interaction, $\hbar \Omega \approx \Delta/2$ (meV)

Tunneling time, $\tau^*$ (ps)

$\tau^* = 0.05$ ps
$\tau^* = 0.15$ ps
$\tau^* = 0.25$ ps
$\tau^* = 0.35$ ps
$\tau^* = 0.45$ ps

Dephasing time, $\tau_\parallel$ (meV)

Figure 2-2: Typical values of Kazarinov-Suris tunneling time $\tau^*$ at resonance ($\omega_{12} = 0$ in equation 2.27). Note that on resonance, the anticrossing $\Delta$ between two delocalized subbands and the interaction $V = \hbar \Omega$ between the corresponding localized states is approximately related by $\Delta \approx 2\hbar \Omega$. 38
3-level system are

\[
\begin{align*}
-\frac{n_1}{\tau_{13}^*} + \frac{n_2}{\tau_{21}} + \frac{n_3}{\tau_{31}} + \frac{n_3}{\tau_{13}^*} &= 0 \text{ (subband 1)} \quad (2.28) \\
-\frac{n_2}{\tau_{21}} + \frac{n_3}{\tau_{32}} &= 0 \text{ (subband 2)} \quad (2.29) \\
\frac{n_1}{\tau_{13}^*} - \frac{n_3}{\tau_{13}^*} - \frac{n_3}{\tau_{32}} - \frac{n_3}{\tau_{31}} &= 0 \text{ (subband 3)} \quad (2.30)
\end{align*}
\]

Only two of these equations are independent, and so the rate equations must be solved by also specifying the normalization condition

\[
n_1 + n_2 + n_3 = n \quad (2.31)
\]

where \( n \) is the total electron population per QCL period.

Below threshold, the rate equations form an algebraic linear system, which is easily solved. For example, the population inversion is

\[
\Delta n = n_3 - n_2 = n \frac{\frac{1}{\tau_{13}^*} \left( 1 - \frac{\tau_{31}}{\tau_{32}} \right)}{\frac{1}{\tau_{13}^*} \left( 2 + \frac{\tau_{31}}{\tau_{32}} \right) + \frac{1}{\tau_{31}} + \frac{1}{\tau_{32}}} \quad (2.32)
\]

and the current density is given by

\[
J = en \frac{\frac{1}{\tau_{13}^*} \left( \frac{1}{\tau_{31}} + \frac{1}{\tau_{32}} \right)}{\frac{1}{\tau_{13}^*} \left( 2 + \frac{\tau_{31}}{\tau_{32}} \right) + \frac{1}{\tau_{31}} + \frac{1}{\tau_{32}}} \quad (2.33)
\]
If equation 2.27 is substituted for $\tau_{13}^*$, then equation 2.33 can be easily demonstrated to be equivalent to Kumar’s equation 2.87 in ref. [33].

**Above threshold**

The analysis of the 3-level laser above threshold is more involved, as the optical intensity in the cavity is needed in addition to the subband populations. Equivalently, one may solve instead for the simulated emission time, as the two are related by

$$\frac{1}{\tau_{st}} = \sigma_{32} S$$

where $S$ is the photon flux, and $\sigma_{32}$ is the gain cross-section for the 3-2 lasing transition. During lasing, the laser gain $g$ clamps to a gain threshold $g_{th}$ determined by the laser’s modal losses. Assuming that all gain comes from the 3-2 transition, then the rate equations must also satisfy the addition restraint that

$$g = \frac{\sigma_{32}}{l} (n_3 - n_2) = \frac{\alpha_w + \alpha_m}{\Gamma} = g_{th}$$

where $l$ is the length of one QCL period, $\Gamma$ is the modal confinement, $\alpha_w$ is the waveguide loss, and $\alpha_m$ is mirror loss. In this simple situation, the population inversion itself clamps to some constant value $\Delta n_{th} = g_{th} l / \sigma_{32}$.

Neglecting spontaneous emission effects, the effective rates equations modified for stimulated emission is

$$-\frac{n_1}{\tau_{13}^*} + \frac{n_2}{\tau_{21}} + \frac{n_3}{\tau_{31}} + \frac{n_3}{\tau_{13}^*} = 0 \text{ (subband 1) (2.36)}$$

$$-\frac{n_2}{\tau_{21}} - \frac{n_2}{\tau_{st}} + \frac{n_3}{\tau_{32}} + \frac{n_3}{\tau_{st}} = 0 \text{ (subband 2) (2.37)}$$

$$\frac{n_1}{\tau_{13}^*} + \frac{n_2}{\tau_{st}} - \frac{n_3}{\tau_{13}^*} - \frac{n_3}{\tau_{32}} - \frac{n_3}{\tau_{31}} - \frac{n_3}{\tau_{st}} = 0 \text{ (subband 3) (2.38)}$$

In contrast to the case below threshold, these equations are nonlinear, and so their
solution is more involved. The equation for subband 2 is used to write

\[ n_2 = \left( \frac{1}{\tau_{32}} + \frac{1}{\tau_{st}} \right) n_3 \]

\[ = Rn_3 = \begin{cases} \frac{n_3 \tau_{21}}{\tau_{31}} & \text{when } \tau_{st} = \infty \\ n_3 & \text{when } \tau_{st} = 0 \end{cases} \]  

(2.39)

where the variable \( \tau_{21}/\tau_{32} < R < 1 \) is a measure of stimulated emission. This is substituted into the equation for subband 1 to write

\[ n_1 = n_3 \tau_{13}^* \left( \frac{R}{\tau_{21}} + \frac{1}{\tau_{31}} + \frac{1}{\tau_{13}^*} \right) \]  

(2.40)

Using the normalization condition 2.31, the upper level population is derived to be

\[ n_3 = \frac{n}{\tau_{13}^* \left( \frac{R}{\tau_{21}} + \frac{1}{\tau_{31}} \right) + R + 2} \]  

(2.41)

and the population inversion is hence

\[ n_3 - n_2 = (1 - R) n_3 \]

\[ = \frac{n (1 - R)}{\tau_{13}^* \left( \frac{R}{\tau_{21}} + \frac{1}{\tau_{31}} \right) + R + 2} = \Delta n_{th} \]  

(2.42)

This can be rearranged to express \( R \) as a function of the threshold. This yields

\[ R = \frac{1 - \frac{\Delta n_{th}}{n} \left( 2 + \frac{\tau_{13}^*}{\tau_{31}} \right)}{1 + \frac{\Delta n_{th}}{n} \left( 1 + \frac{\tau_{13}^*}{\tau_{21}} \right)} \]  

(2.43)
The total current density is thus

\[ J = e \left( \frac{n_3}{\tau_{31}} + \frac{n_2}{\tau_{21}} \right) \]

\[ = en_3 \left( \frac{1}{\tau_{31}} + \frac{R}{\tau_{21}} \right) \]

\[ = en \frac{1}{\tau_{13}} \left( \frac{1}{\tau_{31}} + \frac{R}{\tau_{21}} \right) \]

\[ = \frac{1}{\tau_{13}} \left( 2 + R + \frac{1}{\tau_{31}} + \frac{R}{\tau_{21}} \right) \]

(2.44)

Note that equations 2.33 and 2.44 agree when \( R = \tau_{21}/\tau_{32} \) (corresponding to \( \tau_{st} = \infty \), no lasing). Further algebraic manipulation of equation 2.44 shows it to be equivalent to Kumar’s equation 2.87 in ref. [33].

The stimulated emission rate can furthermore be derived from \( R \). Solving for \( 1/\tau_{st} \) from the definition of \( R \), and using equation 2.43 yields

\[ \frac{\Delta n_{th}}{\tau_{st}} = \frac{1}{\tau_{21}} \left( 2 + \frac{\tau_{13}}{\tau_{31}} \right) + \frac{1}{\tau_{32}} \left( 1 + \frac{\tau_{13}}{\tau_{21}} \right) \left[ n_1 \frac{1}{\tau_{13}} \left( 2 + \frac{\tau_{21}}{\tau_{32}} \right) + \frac{1}{\tau_{31}} + \frac{1}{\tau_{32}} \right] - \Delta n_{th} \]

(2.45)

The first term in the square bracket is the unclamped population inversion specified by equation 2.32. Therefore, equation 2.45 may be rearranged as

\[ \frac{1}{\tau_{st}} = \frac{1}{\tau_{21}} \left( 2 + \frac{\tau_{13}}{\tau_{31}} \right) + \frac{1}{\tau_{32}} \left( 1 + \frac{\tau_{13}}{\tau_{21}} \right) \left[ \frac{\Delta n_{S=0}}{\Delta n_{th}} - 1 \right] \]

(2.46)

Since \( \frac{\Delta n_{S=0}}{\Delta n_{th}} = \frac{g_{S=0}}{g_{th}} \) and \( 1/\tau_{st} \propto S \), equation 2.46 has an appealing physical interpretation: the output power output is driven by the unsaturated gain above threshold.

\[ S \propto \left( \frac{g_{S=0}}{g_{th}} - 1 \right) \]

(2.47)

2.2.3 Second-order tunneling

The original KS model contains some unphysical behavior. The classic example is the case of a simple superlattice consisting of identical quantum wells. By symmetry,
Figure 2-4: Schematic illustration of second-order tunneling. Because electrons tunnel at constant energy, not momentum, all electrons in the higher energy subband can tunnel to the lower energy subband, but only hot electrons in the lower energy subband can tunnel to the higher energy subband.

every well must contain an identical electron distribution, and thus the KS models predicts zero current regardless of bias.

The origins of this unphysical behavior is the implicit assumption of the KS model that electrons do not change in-plane momentum $\mathbf{k}$ during tunneling. In actuality, scattering by interface roughness, impurities, alloy inhomogeneity, and so forth do cause momentum transfer. Wacker, [35] and later Willenberg [36] and Terazzi, [37] have shown that the inclusion of scattering induced higher-order tunneling terms (“second-order”) reveals that tunneling is better described as a constant energy ($E$) process instead of a constant $\mathbf{k}$ process. It is forbidden when no final states of equivalent energy exist (see figure 2-4). The key implication is that out of resonance, the tunneling between two subbands 1 and 2 is not symmetric ($\tau_{12}^* \neq \tau_{21}^*$).

Assuming that the electrons in each subband are Boltzmann distributed, Terazzi essentially amended equation 2.27 to become.

$$\frac{2\Omega^2\tau_{||}}{1 + \omega_{12}^2\tau_{||}^2} \left( \rho_{11}e^{-E'_{21}/k_B T_1} - \rho_{22}e^{-E'_{12}/k_B T_2} \right)$$

(2.48)

where $E'_{ij} = \min(0, E_i - E_j)$. The exponential factors reflects that out of resonance, all electrons in the higher energy subband can tunnel to the lower energy subband, but only the hot tail of the lower energy subband can tunnel into the higher energy.
subband. This refined expression possess nonzero current in the simple superlattice, and it easy to verify that it also satisfies detailed balance in equilibrium (when \( \rho_{22}/\rho_{11} = e^{-E_{21}k_B T} \)).

For typical THz QCLs, however, the differences between equations 2.27 and 2.48 are not so different, especially at high temperatures (\( k_B T \gg E_{21} \)). In this thesis, the more accurate second-order tunneling model of equation 2.48 will be used in numerical calculations (see chapter 3), and the original KS model will be used in analytical calculations.

In addition to transfer of population due to tunneling, the transfer of kinetic energy must also be considered for the calculation of subband temperatures later. The main results for second-order tunneling are summarized below. For tunneling from subband \( a \) to subband \( b \), the population transfer rate is

\[
W^* (a \rightarrow b) = \frac{2\Omega_{ab}^2 \gamma_{ab}}{1 + \omega_{ab,2}^2 \pi \hbar^2} \int_{E_{ba}}^\infty dE f_a (E) \\
= \frac{2\Omega_{ab}^2 \gamma_{ab}}{1 + \omega_{ab,2}^2 \pi \hbar^2} m_e \pi \hbar^2 \int_{E_{ba}}^\infty dE f_a (E)
\]  

(2.49)

When an electron with kinetic energy \( E \) tunnels out of the source subband \( a \), \( a \) loses \( E \) of kinetic energy. The cooling rate of \( a \) is hence given by equation 2.49, but with a factor of \( E \) in the integrand.

\[
R^*_c (a \rightarrow b) = \frac{2\Omega_{ab}^2 \gamma_{ab}}{1 + \omega_{ab,2}^2 \pi \hbar^2} n_a \pi \hbar^2 \int_{E_{ba}}^\infty (k_B T_a + E_{ba}') \exp \left( -\frac{E_{ba}'}{k_B T_a} \right)
\]  

(2.50)

On the other hand, the destination subband \( b \) gains an amount of energy equation to \( E - E_{ba} \), which can be more (less) than \( E \) if the destination subband is lower (higher) in energy than the initial subband. The heating rate of \( b \) is thus

\[
R^*_h (a \rightarrow b) = \frac{2\Omega_{ab}^2 \gamma_{ab}}{1 + \omega_{ab,2}^2 \pi \hbar^2} n_a \pi \hbar^2 \int_{E_{ba}}^\infty (E - E_{ba}) \exp \left( -\frac{E_{ba}'}{k_B T_a} \right)
\]  

(2.51)
2.2.4 The dephasing time

A crucial determinant of the tunneling is the dephasing time $\tau_{||ab}$. In principle, every momentum state has a different rate of dephasing, but this complicates the analysis of tunneling. This work adopts a single dephasing time for every subband pair to represent tunneling of electrons of all momenta. This is estimated as

$$\frac{1}{\tau_{||ab}} = \frac{1}{2\tau_a} + \frac{1}{2\tau_b} + \frac{1}{T^*}$$

(2.52)

where $\tau_a$ and $\tau_b$ are the ensemble average lifetimes of the two subbands. These terms represent dephasing due to intersubband transitions. The last term $T^*$ represents dephasing due to intrasubband scattering (a so-called “pure dephasing” term). Calculation of this term is complicated by the need to consider correlations between the intrasubband scattering in the two subbands. \[37\] Simply adding together the intrasubband scattering rates for $a$ and $b$ yields excessively fast decoherence times (fs scale). For the most part, $T^*$ is treated as a fitting parameter in this work.

2.3 Single-body intersubband scattering

When quantum wells are strongly coupled, the eigenstates of the coupled wells are good descriptors of the quantum system. Electron transport can then be described by the act of scattering between eigenstate subbands. This section details the calculation of single body scattering rates between subbands from Fermi’s golden rule. The usual procedure in the literature involves calculating expressions for each scattering mechanism separately (see, for example \[37\]–\[39\]). Instead of this approach, this section derives a single expression that describes all single-body scattering rates. Further simplification is achieved through assuming Boltzmann distributions in each subband. Complementary to the scattering rates describing transfer of population, expressions are also given for the transfer of kinetic energy between subbands. This enables us to calculate not only the subband populations, but also the subband temperatures (see chapter \[3\]. \[40\]
The single body scattering mechanisms included in this work are longitudinal optical (LO) phonon interactions, impurity scattering, and interface roughness scattering. LO phonon scattering is inelastic, whereas impurity and interface roughness scattering are (elastic). The transitions between a pair of subbands due to these mechanisms are illustrated in the $k$-space diagram of figure 2-5.

2.3.1 Definitions of the rates

For an electron in subband $a$ with in-plane wavevector $\overline{k}_a$, scattering to another state $b\overline{k}_b$ will be dictated by Fermi’s golden rule.

$$W (a\overline{k}_a \rightarrow b\overline{k}_a) = \frac{2\pi}{\hbar} |U (q)|^2 \delta \left( E_b (\overline{k}_b) - E_a (\overline{k}_a) \pm \Delta \right)$$  

(2.53)

where $E_i (\overline{k}) = E_i + \frac{\hbar^2 k^2}{2m_e}$, $E_i$ is the energy of the subband edge, and $\Delta$ is the energy gained or lost due to the scatterer (relevant in the case of inelastic phonon scattering). $|U (q)|^2$ is the matrix element of the interaction. Assuming isotropy and uniformity in the in-plane direction, $|U (q)|^2$ can be assumed to be a function of only the exchanged momentum $\overline{q} = \overline{k}_b - \overline{k}_a$.

As an electron transfers a origin subband $a$ to a destination subband $b$, 3 changes to the subband statistics occurs.
Figure 2-6: Changes in subband statistics as an electron scatters from subband $a$ to subband $b$. The initial subband $a$ loses one electron (pop. -1) and the final subband $a$ gains one electron (pop. +1). Subband $a$ also loses $\hbar^2 k_a^2 / 2m_e$ of kinetic energy (KE), and subband $b$ gains $\hbar^2 k_b^2 / 2m_e$ of KE.

- Subband $a$ loses (and subband $b$ gains) an electron.

- Subband $a$ loses kinetic energy (KE); it cools down.

- Subband $b$ gains kinetic energy (KE); it heats up.

These processes are illustrated in figure 2-6. Consequently, there are three rates of interest. These are the transfer rate,

$$ W (a \rightarrow b) = \frac{2}{A} \sum_{\bar{k}_a} \sum_{\bar{k}_b} W (a \bar{k}_a \rightarrow b \bar{k}_a) f_a (\bar{k}_a) (1 - f (\bar{k}_b)) \quad (2.54) $$

the cooling rate of the initial subband $a$,

$$ R_c (a \rightarrow b) = \frac{2}{A} \sum_{\bar{k}_a} \sum_{\bar{k}_b} \left( \frac{\hbar^2 k_a^2}{2m_e} \right) W (a \bar{k}_a \rightarrow b \bar{k}_a) f_a (\bar{k}_a) (1 - f (\bar{k}_b)) \quad (2.55) $$

and the heating rate of the final subband $b$,

$$ R_h (a \rightarrow b) = \frac{2}{A} \sum_{\bar{k}_a} \sum_{\bar{k}_b} \left( \frac{\hbar^2 k_b^2}{2m_e} \right) W (a \bar{k}_a \rightarrow b \bar{k}_a) f_a (\bar{k}_a) (1 - f (\bar{k}_b)) \quad (2.56) $$

In the above, the prefactor of 2 accounts for spin.
2.3.2 Transfer rates

Under the assumption of negligible state-blocking, and using Fermi’s golden rule, the lifetime of a given state $a\bar{k}_a$ is given by

$$W (a\bar{k}_a \to b) = \sum_{\bar{k}_b} \frac{2\pi}{\hbar^2} |U(q)|^2 \delta (E_b(\bar{k}_b) - E_a(\bar{k}_a) \pm \Delta)$$  \hspace{1cm} (2.57)

Most sources in the literature complete the integral in this form (see for example [39]). Instead of doing this, a change of variables is performed in order to complete the sum over $q$ instead (integration in $q$ is also employed in [41]). For computing the lifetime of a state $a\bar{k}_a$, this confers no advantage, but it becomes useful later when performing an ensemble average over the subband distribution. This yields

$$W (a\bar{k}_a \to b) = \sum_{q} \frac{2\pi}{\hbar^2} |U(q)|^2 \delta (E_b(\bar{k}_b) - E_a(\bar{k}_a) \pm \Delta)$$

$$= \frac{A}{2\pi \hbar} \int q dq |U(q)|^2 \int d\theta \delta (E_b(\bar{k}_b) - E_a(\bar{k}_a) \pm \Delta)$$  \hspace{1cm} (2.58)

where cylindrical coordinates have been deployed. The angular integral may be completed analytically.

$$\int d\theta \delta (E_b(\bar{k}_b) - E_a(\bar{k}_a) \pm \Delta)$$

$$= \int_{-\pi}^{\pi} d\theta \delta \left( E_b - E_a \pm \Delta + \frac{\hbar^2}{2m_e} (q^2 + 2k_a q \cos \theta) \right)$$

$$= \int_{-\pi}^{\pi} d\theta \delta \left( E_b - E_a \pm \Delta + \epsilon_q + 2\sqrt{\epsilon_a \epsilon_q \cos \theta} \right)$$  \hspace{1cm} (2.59)

$$= \frac{2 \Theta (4\epsilon_a \epsilon_q > (E_b - E_a \pm \Delta + \epsilon_q)^2)}{\sqrt{4\epsilon_a \epsilon_q - (E_b - E_a \pm \Delta + \epsilon_q)^2}}$$

48
where $\Theta$ is the unit step function, $\epsilon_a = \frac{\hbar^2 k^2}{2m_e}$ and $\epsilon_q = \frac{\hbar^2 q^2}{2m_e}$. Next, an average is taken over the initial distribution

$$W(a \rightarrow b) = \frac{A}{2\pi \hbar} \int q dq |U(q)|^2 2 \int \frac{d^2 \vec{k}_a}{(2\pi)^2} f_a(\vec{k}_a) \frac{2\Theta(4\epsilon_a \epsilon_q > (E_b - E_a \pm \Delta + \epsilon_q)^2)}{\sqrt{4\epsilon_a \epsilon_q - (E_b - E_a \pm \Delta + \epsilon_q)^2}}$$

Using the assumption that subband distributions $f_a(\vec{k}_a)$ are Boltzmannian, the integral in $\vec{k}_a$ can be completed analytically. Converting to an integral over energy yields

$$2 \int \frac{d^2 \vec{k}_a}{(2\pi)^2} f_a(\vec{k}_a) \frac{2\Theta(4\epsilon_a \epsilon_q > (E_b - E_a \pm \Delta + \epsilon_q)^2)}{\sqrt{4\epsilon_a \epsilon_q - (E_b - E_a \pm \Delta + \epsilon_q)^2}}$$

where a characteristic energy $\epsilon'_a(q) = \frac{(E_b - E_a \pm \Delta + \epsilon_q)^2}{4\epsilon_q}$ has been defined. Inserting this into equation (2.60) yields our desired result.

$$W_a(a \rightarrow b) = n_a \frac{1}{h^2} \sqrt{\frac{m_e}{2\pi k_B T_a}} \int dq [A|U(q)|^2] \exp \left( -\frac{\epsilon'_a(q)}{k_B T_a} \right)$$

The advantage of assuming Boltzmann statistics and neglecting state blocking is that the calculation of the carrier transfer rate is reduced to a single integral. Equation (2.62) describes all single-body scattering mechanisms, which differ only in the choice of matrix element $U(q)$.
2.3.3 Cooling rate

The calculation of the cooling rate proceeds in the same fashion as the transfer rate up to equation [2.60]. There, an addition factor of $\epsilon_a = \frac{\hbar^2 k_b^2}{2m_e}$ is needed in the integrand. Therefore, the average over the distribution is modified to

$$\frac{m_e n_a}{\pi \hbar^2} \left( \frac{2 \pi \hbar^2}{m_e k_B T_a} \right) \int d\epsilon_a \epsilon_a \exp \left( -\frac{\epsilon_a}{k_B T_a} \right) \frac{2 \Theta \left( 4 \epsilon_a \epsilon_q > (E_b - E_a \pm \Delta + \epsilon_q)^2 \right)}{\sqrt{4 \epsilon_a \epsilon_q - (E_b - E_a \pm \Delta + \epsilon_q)^2}}$$

and therefore, equation [2.55] becomes

$$R_a (a \rightarrow b) = n_a \frac{1}{\hbar^2} \sqrt{\frac{m_e}{2 \pi k_B T_a}} \int dq \left( \frac{1}{2} k_B T_a + \epsilon_0 q \right) \left[ A |U(q)|^2 \right] \exp \left( -\frac{\epsilon_0 q}{k_B T_a} \right)$$

2.3.4 Heating rate

Energy conservation implies

$$\frac{\hbar^2 k_b^2}{2m_e} = \frac{\hbar^2 k_a^2}{2m_e} - (E_b - E_a \pm \Delta)$$

(2.65)

Therefore, from the definition of $R_h$ (equation [2.56]), the heating rate is given by

$$R_h (a \rightarrow b) = \frac{1}{A} \sum_{k_a} \sum_{k_b} \left( \frac{\hbar^2 k_b^2}{2m_e} - (E_b - E_a \pm \Delta) \right) W \left( a \bar{k}_a \rightarrow b \bar{k}_a \right) W \left( k_a \rightarrow b \bar{k}_a \right) \left( 1 - f \left( \bar{k}_b \right) \right)$$

$$= \frac{1}{A} \sum_{k_a} \sum_{k_b} \left( \frac{\hbar^2 k_a^2}{2m_e} - (E_b - E_a \pm \Delta) \right) W \left( a \bar{k}_a \rightarrow b \bar{k}_a \right) \left( 1 - f \left( \bar{k}_b \right) \right)$$

$$= R_c (a \rightarrow b) - (E_b - E_a \pm \Delta) W (a \rightarrow b)$$

(2.66)
where the exchanged energy \( \Delta \) is assumed to be a constant. This is approximately true for LO phonons, the dominate inelastic scatterers (but will fail for acoustic phonons). Thus, calculating the kinetic energy gain of subband \( b \) requires a trivial amount of work once the transfer and cooling rates are known.

### 2.3.5 Matrix elements for single body scattering

Given the computational framework given in section 2.3, what remains is to specify the matrix elements \( |U(q)|^2 \). This is the subject of this section.

This section uses box normalization. Wavefunctions are assumed to have the form

\[
\psi(\bar{r}) = (\phi_{cb}(z) u_{cb}(\bar{r}) + \phi_{lh}(z) u_{cb}(\bar{r}) + \phi_{so}(z) u_{cb}(\bar{r})) \frac{e^{i\bar{k} \cdot \rho}}{\sqrt{A}}
\]

(2.67)

where \( \phi \) are the band envelopes, and \( u \) are the Bloch amplitudes. Under the assumption of scatterings potentials \( H' \) and envelopes \( \phi \) that are slowly varying with respect to a lattice constant, and using the orthogonality of the Bloch amplitudes, the matrix elements are well approximated by

\[
\langle a\bar{k}_a | H' | b\bar{k}_b \rangle \approx \sum_\beta \int dz \phi_{\beta a}^* e^{-i\bar{k}_a \cdot \rho} H' \phi_{\beta b} e^{i\bar{k}_b \cdot \rho} \frac{1}{\sqrt{A}}
\]

(2.68)

where the summation \( \beta \) runs over the bands \( cb, lh, so \).

Furthermore, the following Fourier transform convention is assumed.

\[
f(\bar{\rho}) = \frac{1}{A} \sum_\bar{q} f(\bar{q}) e^{i\bar{q} \cdot \bar{\rho}} \leftrightarrow f(\bar{\rho}) = \int d^2 \bar{\rho} f(\bar{\rho}) e^{-i\bar{q} \cdot \bar{\rho}}
\]

(2.69)

### Longitudinal optical phonon scattering

This matrix element and its derivation is well documented in the literature (see [38] for example), so it is simply stated here for completeness.

\[
|U(q)|^2 = \frac{e^2 \hbar \omega_{LO}}{4A} \left( \frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_s} \right) \left( N_{LO} + \frac{1}{2} \pm \frac{1}{2} \right) \left[ \frac{F_{abab}(q)}{q} \right]
\]

(2.70)
where the symbols are defined as follows.

- $e > 0$ is the (magnitude of) the electron charge.
- $\hbar \omega_{LO} \sim 36$ meV is the LO phonon energy.
- $\varepsilon_{\infty}$ and $\varepsilon_s$ are the optical and static permittivities of GaAs.
- $N_{LO}$ is the phonon occupation factor (Bose-Einstein form in thermal equilibrium).
- $F_{abab}(q)$ is the Coulombic form factor (see section 3.4.1).
- $A$ is a normalization area.

The exchanged energy is $\Delta = \hbar \omega_{LO}$; the upper sign is taken for phonon emission, while the lower sign is taken for phonon absorption.

In this treatment, the LO phonons are taken to be dispersionless. However, the real LO phonon dispersion of GaAs decreases by $\sim 6$ meV between $\Gamma$-point and the zone edges, which is significant on THz energy scales ($6$ meV $\equiv 1.5$ THz). To answer whether this dispersion is significant, consider the exchanged momentum $q$. Even for intrasubband LO phonon scattering, $q = \sqrt{2m_e \hbar \omega_{LO} / \hbar} \approx 2.5 \times 10^8$ m$^{-1}$. This is very small compared to the extent of the Brillouin zone ($2\pi/a \sim 10^{10}$ m$^{-1}$), so the LO phonon dispersion is not expected to be significant.

**Interface roughness scattering**

The model of Unuma is employed, in which the interface roughness is treated as a scattering mechanism. The perturbing Hamiltonian is

$$H' = \sum_n \Delta_n h_n(\mathbf{r}) \delta(z-z_n) \quad (2.71)$$

where the summation is taken over all interfaces, and location of the $n$-th interface is at $z_n$. $\Delta_n$ is the band discontinuity at the interface $n$ (defined to be positive when the potential on right of interface greater than left of interface, and negative vice versa),

52
and $h_n$ is the interface roughness function (deviation from the interface mean). The matrix element is hence

$$\langle ak | H' | bk \rangle = \sum_{\beta n} \Delta_n \phi_{\beta a} (z_n) \phi_{\beta b} (z_n) \frac{1}{A} \int d^2 \tau h_n (\tau) e^{i \overline{q} \cdot \tau}$$

$$= \sum_{\beta n} (\phi_{\beta a} \phi_{\beta b} \Delta)_{\beta n} \frac{1}{A} \int d^2 \tau h_n (\tau) e^{i \overline{q} \cdot \tau}$$

$$= \sum_n F_{ab} \frac{1}{A} \int d^2 \tau h_n (\tau) e^{i \overline{q} \cdot \tau}$$

(2.72)

where

$$F_{ab} = \langle \phi_{a} \phi_{b} \Delta \rangle_{cb, n} + \langle \phi_{a} \phi_{b} \Delta \rangle_{lh, n} + \langle \phi_{a} \phi_{b} \Delta \rangle_{so, n}$$

(2.73)

and $\overline{q} = \overline{k}_b - \overline{k}_a$ is the exchanged momentum. The squared matrix element required in Fermi’s golden rule is hence

$$|\langle ak | H | bk \rangle|^2$$

$$= \left[ \sum_m F_{mab} \frac{1}{A} \int d^2 \overline{\rho} h_m (\overline{\rho}) e^{-i \overline{q} \cdot \overline{\rho}} \right] \left[ \sum_n F_{nab} \frac{1}{A} \int d^2 \overline{\rho} h_n (\overline{\rho}) e^{i \overline{q} \cdot \overline{\rho}} \right]$$

$$= \sum_{mn} F_{mab} F_{nab} \frac{1}{A^2} \int d^2 \overline{\rho} \int d^2 \overline{\rho}' h_m (\overline{\rho}) h_n (\overline{\rho}') e^{i \overline{q} \cdot (\overline{\rho} - \overline{\rho}')}$$

$$= \frac{1}{A} \sum_{mn} F_{mab} F_{nab} \left[ \int d^2 \overline{\rho}_0 \left( \frac{1}{A} \int d^2 \overline{\rho}' h_m (\overline{\rho}') h_n (\overline{\rho}_0 + \overline{\rho}') \right) e^{i \overline{q} \cdot \overline{\rho}_0} \right]$$

$$= \frac{1}{A} \sum_{mn} F_{mab} F_{nab} \left[ \int d^2 \overline{\rho}_0 R_{mn} (\overline{\rho}_0) e^{i \overline{q} \cdot \overline{\rho}_0} \right]$$

(2.74)

where $R_{mn} (\overline{\rho})$ is the cross-correlation between two interfaces (or autocorrelation when $m = n$); the Fourier transform of this quantity is what determines scattering. For simplicity, all interfaces are assumed to have the same autocorrelation, given by the exponential form

$$R (\overline{\rho}) = h^2 \exp \left( - \frac{\overline{\rho}}{\Lambda} \right)$$

(2.75)

which has the Fourier transform

$$R (\overline{q}) = \frac{2 \pi h^2 \Lambda^2}{(1 + q^2 \Lambda^2)^{3/2}}$$

(2.76)
where $h$ is the interface roughness height (typically $\sim 0.3$ nm), and $\Lambda$ is the auto-correlation length (typically 5 nm). Terazzi’s example [37] is followed for the cross-correlation terms, in which they are assume to have the form

$$R_{mn} (\bar{p}) = \sqrt{R_{mm} (\bar{p}) R_{nn} (\bar{p})} \exp \left( - \frac{|z_m - z_n|}{L_v} \right)$$

$$= R (\bar{p}) c_{mn}$$

(2.77)

where $L_v$ is a vertical correlation length (typically 1.5 nm). Therefore, the matrix element for interface roughness scattering is

$$|U (q)|^2 = R (q) \frac{1}{A} \sum_{mn} F_{m}^{ab} F_{n}^{ab} c_{mn}$$

(2.78)

**Impurity scattering**

Given a (static) impurity interaction $Y (\bar{r} - \bar{r}_n)$ for a set of dopant locations $\bar{r}_n$, the interaction Hamiltonian is

$$H' = \sum_n Y (\bar{r} - \bar{r}_n)$$

(2.79)

The matrix element is

$$\langle a_k | H' | b_k \rangle = \frac{1}{A} \sum_n \int d^2 \rho Y_{j,ab} (z_n, \bar{r} - \bar{r}_n) e^{i \bar{r} \cdot \rho}$$

$$= \frac{1}{A} \sum_n Y_{ab} (z_n, \bar{q}) e^{i \bar{q} \cdot \bar{r}_n}$$

(2.80)

where $Y_{ab} (z_n, \bar{q}) = \sum_j Y_{j,ab} (z_n, \bar{q})$, $j = cb, lh, so$. The matrix element squared is therefore

$$|\langle a_k | H' | b_k \rangle|^2 = \frac{1}{A^2} \sum_{mn} Y_{ab}^* (z_m, \bar{q}) Y_{ab} (z_n, \bar{q}) e^{i \bar{q} \cdot (\bar{r}_n - \bar{r}_m)}$$

$$= \frac{1}{A^2} \sum_n |Y_{ab} (z_n, \bar{q})|^2 + \frac{1}{A^2} \sum_{m \neq n} \sum_{m \neq n} Y_{ab}^* (z_m, \bar{q}) Y_{ab} (z_n, \bar{q}) e^{i \bar{q} \cdot (\bar{r}_n - \bar{r}_m)}$$

(2.81)
Given that the impurity locations are not exactly known, the method of impurity averaging is employed. The sums over \( m, n \) are replaced by integrals over the doping distribution \( N_D(\tau) = N_D(z) \), which may be interpreted as a probability distribution in space for the dopants. The first term is easy, becoming

\[
\frac{1}{A^2} \sum_n |Y_{ab}(z_n, \bar{\tau})|^2 \\
\approx \frac{1}{A^2} \int d^2 \rho_n \int d z_n N_D(z_n) |Y_{ab}(z_n, \bar{\tau})|^2
\tag{2.82}
\]

\[
= \frac{1}{A} \int d z N_D(z) |Y_{ab}(z, \bar{\tau})|^2
\]

The second term becomes

\[
\frac{1}{A^2} \sum_m \sum_{m \neq n} Y_{ab}^*(z_m, \bar{\tau}) Y_{ab}(z_n, \bar{\tau}) e^{i \bar{\tau}(\rho_n - \rho_m)} \\
\approx \frac{1}{A^2} \int d z_m \int d z_n N_D(z_m) N_D(z_n) Y_{ab}^*(z_m, \bar{\tau}) Y_{ab}(z_n, \bar{\tau}) \int d^2 \rho_m \int d^2 \rho_n e^{i \bar{\tau}(\rho_n - \rho_m)} \\
= \left| \int d z N_D(z) Y_{ab}(z, 0) \right|^2 \delta_{\bar{\tau}=0}
\tag{2.83}
\]

The second term is non-zero only when \( \bar{\tau} = 0 \); as impurity scattering is elastic, no momentum exchange implies no scattering at all. As such, the second term appears to be related to band-bending rather than impurity scattering. Therefore, the matrix element is

\[
|U(q)|^2 = \frac{1}{A} \int d z N_D(z) |Y_{ab}(z, \tau)|^2
\tag{2.84}
\]

The choice of impurity interaction \( Y \) depends on the screening model used.

### 2.3.6 Intersubband gain

Interactions with the optical field may also be regarded as a form of single-body interaction. However, the energy-conserving \( \delta \)-function of Fermi’s golden rule yields a divergent expression. The \( \delta \)-function is normally phenomenologically replaced by a homogeneously broadened Lorentzian.
Alternative, the gain can be calculated using the density matrix approach. \[33\] The intersubband permittivity between two subbands 1 and 2 in the relaxation time approximation is given by

\[
\chi(\omega) = -\frac{e^2}{\varepsilon_0} |z_{12}|^2 \frac{1}{l} (n_2 - n_1) \left[ \frac{1}{E_{21} + \hbar \omega + i\hbar \Gamma} + \frac{1}{E_{21} - \hbar \omega - i\hbar \Gamma} \right]
\] (2.85)

where \(l\) is the QCL module length and \(n_2 - n_1\) is the population difference in units of inverse area. The permittivity is then related to the material gain by

\[
g(\omega) = \frac{\omega}{cn} \text{Im} \{\chi(\omega)\}
\] (2.86)

Note that the use of the rotating wave approximation is possible, but unnecessary. The full gain spectrum is formed by the sum over all subband pairs (absorption to continuum states is neglected).

### 2.4 Two-body intersubband scattering

In addition to scattering off phonons, impurities, and interfaces, electrons can interact with one another (e-e interactions). One may distinguish between three types of e-e interactions interactions. First is the mean-field or Hartree interactions, which corresponds to band-bending induced by the real-space location of the electrons; this is typically handled as a modification to the eigenstates through self-consistent solution of the coupled Poisson and Schrodinger equations. The second is the pairwise interactions of electrons with one another. The third is a many-body effect, the interactions of electrons with plasmons, which are the collective oscillations of a free electron gas. These latter two interactions are true scattering mechanisms, in that unlike the first they can cannot be folded into the stationary diagonalization of the unperturbed Hamiltonian. This thesis treats only pair-wise interactions, under the presumption that plasmon effects are less important due to the low doping of THz QCLs. For a pair of subbands, possible pairwise interactions between two electrons are depicted in figure 2-7.
For arbitrary electron distributions, e-e scattering is very difficult to treat. In NEGF, electron-electron scattering is handled through inclusion of the $GW$ self-energy, which requires a three-dimensional convolution integral. The $GW$ self-energy has been calculated only under some very strong assumptions (e.g., assuming scattering off a uniform background gas, or using the plasmon-pole approximation). In ensemble Monte Carlo simulations, pair-wise e-e scattering is commonly included, but at significant computational expense through a three-dimensional integral. The Monte Carlo process itself can be regarded as another degree of

\[ \Sigma_{ab}(\mathbf{k}, E) = i\hbar \int \frac{d^2k'}{(2\pi)^2} \int \frac{dE'}{2\pi\hbar} \sum_{cd} G_{cd}^{<}(\mathbf{k}', E') W_{acdb}^{<}(\mathbf{k} - \mathbf{k}', E - E') \]

A full discussion of NEGF is far, far beyond the scope of this thesis. Ref. provides a useful review.
integration, for even greater computational expense. In ref. [49], Slingerland uses the Monte Carlo pair-wise expression together with the assumption of thermally distributed subbands. The ensemble average over the initial subband requires yet another integral, bringing the total again to four.

This section, derives more efficient expressions for intersubband e-e scattering in THz QCL transport. By using the assumption of Boltzmannian subbands, and neglecting state-blocking, an expression requiring only a single integral to complete is derived. The reduction from 3D/4D to 1D integrals confers significant computational savings.

Note that the assumption of Boltzmannian subbands inherently carries an assumption of ultrafast intrasubband e-e scattering. Therefore, the interest here is primarily in examining intersubband transport.

2.4.1 Transfer rate

Using Fermi’s golden rule, the general expression for scattering of an electron from a state $i\bar{k}_i$ to a state $j\bar{k}_j$ through interaction with a partner electron scattering from subband $a$ to subband $b$ is

$$W (i\bar{k}_i, a \rightarrow j\bar{k}_j, b) = \sum_{\bar{k}_a} \sum_{\bar{k}_b} \frac{2\pi}{\hbar} \left| \frac{1}{A} W_{ijab} (q, \omega) \right|^2 \delta_{\bar{k}_a + \bar{k}_b = \bar{k}_j + \bar{k}_i} \delta (E_i (k_i) + E_a (k_a) - E_j (k_j) - E_b (k_b)) f_a (k_a) (1 - f_j (k_j)) (1 - f_b (k_b))$$

(2.87)

where $W_{ijab} (q, \omega)$ is the Fourier transform of the screened coulombic matrix element (units of energy-area), $\bar{q} = \bar{k}_j - \bar{k}_i$ is the exchanged momentum, and $\hbar \omega = E_j + \frac{\hbar^2 k_j^2}{2m_e} - E_i - \frac{\hbar^2 k_i^2}{2m_e}$ is the exchanged energy. The sum in $\bar{k}_b$ can be eliminated using the Kronecker $\delta$-function. Neglecting state-blocking, and assuming static screening (commonly done for THz QCLs), a sum is taken over the final states through an
integral over $\mathbf{q}$ to yield

$$W (\mathbf{i} \mathbf{k}_i, a \rightarrow j, b) = \frac{2 \pi}{h} \frac{1}{A} \sum_{k_a} f_a (k_a) \frac{1}{A} \sum_{q} |W_{ijab} (q)|^2 \delta \left( E_i (k_i) + E_a (k_a) - E_j (\mathbf{k}_i + \mathbf{q}) - E_b (\mathbf{k}_a - \mathbf{q}) \right)$$

$$= \frac{2 \pi}{h} \frac{1}{A} \sum_{k_a} f_a (k_a) \frac{1}{A} \sum_{q} |W_{ijab} (q)|^2 \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} \mathbf{q} \cdot \mathbf{k}_a \right)$$

(2.88)

A coordinate system oriented along $\mathbf{k}_{ai} = \mathbf{k}_a - \mathbf{k}_i$ is chosen for integration, such that $\mathbf{k}_{ai} \cdot \mathbf{q} = k_{ai} q \cos \theta$. Then the integral in $\mathbf{q}$ becomes

$$\frac{1}{A} \sum_{q} |W_{ijab} (q)|^2 \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} \mathbf{q} \cdot \mathbf{k}_a \right)$$

$$= \frac{1}{(2 \pi)^2} \int q dq |W_{ijab} (q)|^2 \int d\theta \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} q k_{ai} \cos \theta \right)$$

(2.89)

To complete the angular integral, let $u = \cos \theta$. Then

$$\int d\theta \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} q k_{ai} \cos \theta \right)$$

$$= 2 \Theta \left( \left| \frac{\hbar^2 q^2}{m_e} - (E_i + E_a - E_j - E_b) \right| < \frac{\hbar^2}{m_e} q k_{ai} \right)$$

$$\sqrt{\left( \frac{\hbar^2 k_{ai} q}{m_e} \right)^2 - \left( \frac{\hbar^2 q^2}{m_e} - (E_i + E_a - E_j - E_b) \right)^2}$$

$$= \frac{2 m_e}{\hbar^2 q} \Theta \left( |k_{ai}'| < k_{ai} \right)$$

(2.90)

where $k_{ai}'$ is defined by

$$k_{ai}' = q - \frac{m_e}{\hbar^2 q} (E_i + E_a - E_j - E_b)$$

(2.91)
Inserting the results into \(2.88\), the scattering rate is thus given by

\[
W (\mathbf{k}_i, a \rightarrow j, b) = \frac{1}{2\pi} \frac{2m_e}{\hbar^2} \int dq |W_{ijab} (q)|^2 \frac{1}{A} \sum_{k_a} f_a (k_a) \frac{\Theta (|k'_a| < k_{ai})}{\sqrt{k_{ai}^2 - k'^2_{ai}}} \tag{2.92}
\]

To complete the \(\bar{k}_a\) integral, a change of variables to \(\bar{k}_{ai}\) is made.

\[
\frac{1}{A} \sum_{\bar{k}_{ai}} f_a (k_i + \bar{k}_{ai}) \Theta (k'_{ai} < k_{ai}) \frac{\Theta (|k'_a| < k_{ai})}{\sqrt{k_{ai}^2 - k'^2_{ai}}}
\]

\[
= \frac{1}{(2\pi)^2} \frac{n_a}{2} \left( \frac{2\pi \hbar^2}{m_e k_B T_a} \right) \int d\bar{k}_{ai} \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_{ai} + k^2_{ai} + 2k_i \cdot \bar{k}_{ai}) \right) \frac{\Theta (|k'_a| < k_{ai})}{\sqrt{k_{ai}^2 - k'^2_{ai}}} \tag{2.93}
\]

\[
= n_a \left( \frac{\hbar^2}{4\pi m_e k_B T_a} \right) \int_{|k'_a|}^{\infty} \frac{k_{ai} dk_{ai}}{\sqrt{k_{ai}^2 - k'^2_{ai}}} \int d\theta' \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_{ai}^2 + k^2_{ai} + 2k_i k_{ai} \cos \theta') \right)
\]

where \(\theta'\) is angle between \(k_i\) and \(\bar{k}_{ai}\). The angular integral is completed as

\[
\int d\theta' \exp \left( -\frac{\hbar^2}{m_e k_B T_a} k_i k_{ai} \cos \theta' \right) = 2\pi I_0 \left( \frac{\hbar^2 k_i k_{ai}}{m_e k_B T_a} \right) \tag{2.94}
\]

Therefore

\[
\frac{1}{A} \sum_{\bar{k}_{ai}} f_a (k_i + \bar{k}_{ai}) \Theta (k'_{ai} < k_{ai}) \frac{\Theta (|k'_a| < k_{ai})}{\sqrt{k_{ai}^2 - k'^2_{ai}}}
\]

\[
= n_a \left( \frac{\hbar^2}{2m_e k_B T_a} \right) \int_{|k'_a|}^{\infty} \frac{k_{ai} dk_{ai}}{\sqrt{k_{ai}^2 - k'^2_{ai}}} \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_{ai}^2 + k^2_{ai}) \right) I_0 \left( \frac{\hbar^2 k_i k_{ai}}{m_e k_B T_a} \right)
\]

\[
= n_a \left( \frac{\hbar^2}{2m_e k_B T_a} \right) \int_0^{\infty} d\kappa \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_{ai}^2 + \kappa^2 + k'^2_{ai}) \right) I_0 \left( \frac{\hbar^2 k_i \sqrt{\kappa^2 + k'^2_{ai}}}{m_e k_B T_a} \right) \tag{2.95}
\]
where $\kappa^2 = k_{ai}^2 - k_{ai}'^2$. The scattering rate is hence

$$W(i \bar{k}_i, a \rightarrow j, b) = \left( \frac{n_a}{2\pi \hbar k_B T_a} \right) \int dq |W_{ijab}(q)|^2 \int_0^\infty d\kappa \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_i^2 + \kappa^2 + k_{ai}'^2) \right) I_0 \left( \frac{\hbar^2 k_i \sqrt{\kappa^2 + k_{ai}'^2}}{m_e k_B T_a} \right) \quad (2.96)$$

Next, a sum is taken over the distribution of the initial subband, accounting for spin.

$$W(i, a \rightarrow j, b) = 2 \left( \frac{n_a}{2\pi \hbar k_B T_a} \right) \left( \frac{2\pi \hbar^2}{m_e k_B T_i} \right) \frac{1}{2\pi} \int dq |W_{ijab}(q)|^2 \int k_i dk_i \exp \left( -\frac{\hbar^2 k_i^2}{2m_e k_B T_i} \right) \int_0^\infty d\kappa \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_i^2 + \kappa^2 + k_{ai}'^2) \right) I_0 \left( \frac{\hbar^2 k_i \sqrt{\kappa^2 + k_{ai}'^2}}{m_e k_B T_a} \right)$$

$$= \frac{\hbar}{2\pi m_e} \left( \frac{n_a}{k_B T_a} \right) \left( \frac{n_i}{k_B T_i} \right) \int dq |W_{ijab}(q)|^2 \int_0^\infty d\kappa \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_i^2 + \kappa^2) \right) I_0 \left( \frac{\hbar^2 k_i \sqrt{\kappa^2 + k_{ai}'^2}}{m_e k_B T_a} \right)$$

$$= \frac{1}{2\pi \hbar k_B (T_a + T_i)} \int dq |W_{ijab}(q)|^2 \int_0^\infty d\kappa \exp \left( -\frac{\hbar^2 (\kappa^2 + k_{ai}'^2)}{2m_e k_B (T_a + T_i)} \right) \quad (2.97)$$

This can be completed using the analytic integral

$$\int_0^\infty x dx \exp (-ax^2) I_0 (bx) = \frac{1}{2a} \exp \left( \frac{b^2}{4a} \right) \quad (2.98)$$

to arrive at

$$W(i, a \rightarrow j, b) = \frac{1}{2\pi \hbar k_B (T_a + T_i)} \int dq |W_{ijab}(q)|^2 \int_0^\infty d\kappa \exp \left( -\frac{\hbar^2 (\kappa^2 + k_{ai}'^2)}{2m_e k_B (T_a + T_i)} \right) \quad (2.99)$$
Since \( k'_{ai} \) does not depend on \( \kappa \), the integral in \( \kappa \) can be completed easily, yielding.

\[
W(i, a \rightarrow j, b) = n_a n_i \frac{1}{2\hbar^2} \sqrt{\frac{m_e}{2\pi k_B(T_a + T_i)}} \int dq |W_{ijab}(q)|^2 \exp \left( -\frac{\hbar^2 k'^2_{ai}}{2m_e k_B(T_a + T_i)} \right)
\]

(2.100)

Possessing only a single 1-D integral, equation 2.100 is much faster than other approaches in the literature.

### 2.4.2 Cooling rate

Calculating the cooling rate of the initial subband \( i \) starts by returning to equation 2.96. A sum over the initial subband distribution accounting for spin is again made, but this time weighted against the kinetic energy.

\[
R_{ci}(i, a \rightarrow j, b) = \frac{\hbar}{2\pi m_e} \left( \frac{n_a}{k_B T_a} \right) \left( \frac{n_i}{k_B T_i} \right) \int dq |W_{ijab}(q)|^2 \int_0^\infty d\kappa \exp \left( -\frac{\hbar^2}{2m_e k_B T_a T_i} \left( k^2 + k'^2_{ai} \right) \right) \\
\int k_i dk_i \left( \frac{\hbar^2 k_i^2}{2m_e} \right) \exp \left( -\frac{\hbar^2}{2m_e} \left( \frac{1}{k_B T_a} + \frac{1}{k_B T_i} \right) k_i^2 \right) I_0 \left( \frac{\hbar^2 k_i \sqrt{k^2 + k'^2_{ai}}}{m_e k_B T_a} \right)
\]

(2.101)

To continue, the analytic integral

\[
\int_0^\infty x^3 dx \exp \left( -ax^2 \right) I_0(bx)
\]

\[
= \frac{1}{2a} \exp \left( \frac{b^2}{4a} \right) \frac{1}{a} \left( 1 + \frac{b^2}{4a} \right)
\]

(2.102)

can be used to arrive at

\[
R_{ci}(i, a \rightarrow j, b) = n_a n_i \frac{1}{2\hbar^2} \sqrt{\frac{m_e}{2\pi k_B(T_a + T_i)}} \left( \frac{k_B T_a T_i}{T_a + T_i} \right) \int dq |W_{ijab}(q)|^2 \\
\exp \left( -\frac{\hbar^2 k'^2_{ai}}{2m_e k_B(T_a + T_i)} \right) \left( 1 + \frac{1}{2} \left( \frac{T_i}{T_a} \right) + \left( \frac{T_i}{T_a} \right) \frac{\hbar^2 k'^2_{ai}}{2m_e k_B(T_a + T_i)} \right)
\]

(2.103)
Exchanging the indices \(i\) and \(a\), \(j\) and \(b\) yields \(R_{ca}\), the cooling of the partner subband.

### 2.4.3 Heating rate

In contrast to the cooling rate, derivation of the heating rate of the final subband is substantially more complicated. The derivation starts with defining the heating of \(j\) due to a single electron scattering to \(j\) from \(i k_i\).

\[
R_{hj} (i k_i, a \rightarrow j, b) = \frac{2\pi}{\hbar} \sum_{k_a} f_a (k_a) \frac{1}{A} \sum_{q} \frac{\hbar^2 k_i^2}{2m_e} |W_{ijab} (q)|^2 \delta \left( E_i (k_i) + E_a (k_a) - E_j (k_i + q) - E_b (k_a - q) \right)
\]

\[
= \frac{2\pi}{\hbar} \sum_{k_a} f_a (k_a) \frac{1}{A} \sum_{q} \frac{\hbar^2}{2m_e} (k_i^2 + q^2 + 2k_i \cdot q) |W_{ijab} (q)|^2 \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} k_i \cdot q \right)
\]

The term proportional to \(k_i^2\) yields the cooling rate. The terms proportional to \(q^2\) and \(k_i \cdot q\), will be denoted \(R_{hj1}\) and \(R_{hj2}\) respectively. From the derivation of the transfer rate (equation 2.100), clearly

\[
R_{hj1} (i, a \rightarrow j, b) = n_a n_i \frac{1}{2\hbar^2} \sqrt{\frac{m_e}{2\pi k_B (T_a + T_i)}} \int dq |W_{ijab} (q)|^2 \frac{\hbar^2 q^2}{2m_e} \exp \left( -\frac{\hbar^2 k_i^2}{2m_e k_B (T_a + T_i)} \right) \tag{2.105}
\]

In contrast, the derivation of \(R_{hj2}\) is quite involved.

\[
R_{hj2} (i k_i, a \rightarrow j, b) = \frac{2\pi}{\hbar} \sum_{k_a} f_a (k_a) \frac{1}{A} \sum_{q} \frac{\hbar^2}{m_e} k_i \cdot q |W_{ijab} (q)|^2 \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} k_i \cdot q \right) \tag{2.106}
\]

To continue, first isolate the \(q\) integral,
\[
\frac{1}{A} \sum_q \mathbf{k}_i \cdot \bar{q} |W_{ijab}(q)|^2 \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} \bar{q} \cdot \mathbf{k}_{ai} \right) \tag{2.107}
\]

For integration, coordinates aligned with \( \mathbf{k}_{ai} \) are chosen, such that \( \theta \) is the angle between \( \mathbf{k}_{ai} \) and \( \bar{q} \), and \( \theta' \) is the angle between \( \mathbf{k}_i \) and \( \mathbf{k}_{ai} \). This leads us to

\[
\frac{1}{A} \sum_q \mathbf{k}_i \cdot \bar{q} |W_{ijab}(q)|^2 \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} \mathbf{k}_{ai} \mathbf{q} \right) \]

\[= \frac{1}{(2\pi)^3} \int q dq q |W_{ijab}(q)|^2 \int d\theta \cos (\theta - \theta') \]

\[\delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} k_{ai} q \cos \theta \right) \tag{2.108}\]

The angular integral can be completed as

\[
\int d\theta \cos (\theta - \theta') \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} k_{ai} \cos \theta \right)
\]

\[= \int d\theta \cos (\theta - \theta') \delta \left( E_i + E_a - E_j - E_b - \frac{\hbar^2 q^2}{m_e} + \frac{\hbar^2}{m_e} k_{ai} q \cos \theta \right)
\]

\[= \cos (\theta - \theta') \frac{2\Theta \left( \frac{\hbar^2 q^2}{m_e} - (E_i + E_a - E_j - E_b) \right) < \frac{k_{ai} q}{\hbar^2}}{\sqrt{\left( \frac{\hbar^2 k_{ai} q}{m_e} \right)^2 - \left( \frac{k_{ai} q}{m_e} - (E_i + E_a - E_j - E_b) \right)^2}} \tag{2.109}\]

where \( \theta \) is now given by

\[\theta = \cos^{-1} \left( \frac{k_{ai}'}{k_{ai}} \right) \tag{2.110}\]

Plugging the previous four results into equation 2.106 yields

\[
R_{\hbar l 2} (ik_i, a \rightarrow j, b)
\]

\[= \frac{1}{\pi \hbar} \int q dq k_i |W_{ijab}(q)|^2 \frac{1}{A} \sum_{k_a} f_a (k_a) \cos (\theta - \theta') \Theta \left( \frac{|k_{ai}'|}{k_{ai}} \right) \frac{\Theta \left( \frac{k_{ai}'}{k_{ai}} \right)}{\sqrt{k_{ai}^2 - k_{ai}'^2}} \tag{2.111}\]
Next the $k_a$ integral is examined, but through switching to integration in $k_{ai}$.

\[
\frac{1}{A} \sum_{\overline{k}_{ai}} f_a (\overline{k}_i + \overline{k}_{ai}) \cos (\theta - \theta') \frac{\Theta (|k'_{ai}| < k_{ai})}{\sqrt{k_{ai}^2 - k_{ai}'^2}} \nonumber
\]

\[
= n_a \left( \frac{\hbar^2}{4\pi m_e k_B T_a} \right) \int_{k_{ai}'}^\infty \frac{d k_{ai}}{\sqrt{k_{ai}^2 - k_{ai}'^2}} \int d \theta' k_{ai} \cos (\theta - \theta') \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_i^2 + k_{ai}^2 + 2k_i k_{ai} \cos \theta') \right) \nonumber
\]

(2.112)

The quantity $k_{ai} \cos (\theta - \theta')$ expands as

\[
k_{ai} \cos (\theta - \theta')
= k_{ai} (\cos \theta \cos \theta' + \sin \theta \sin \theta')
= k_{ai}' \cos \theta' + \sin \theta' \sqrt{k_{ai}^2 - k_{ai}'^2}
\]

(2.113)

Therefore,

\[
\frac{1}{A} \sum_{\overline{k}_{ai}} f_a (\overline{k}_i + \overline{k}_{ai}) \cos (\theta - \theta') \frac{\Theta (|k'_{ai}| < k_{ai})}{\sqrt{k_{ai}^2 - k_{ai}'^2}}
\]

\[
= n_a \left( \frac{\hbar^2}{4\pi m_e k_B T_a} \right) \int_{k_{ai}'}^\infty dk_{ai} \exp \left( -\frac{\hbar^2}{2m_e k_B T_a} (k_i^2 + k_{ai}^2) \right) \int d \theta' \left( \frac{k_{ai}'}{\sqrt{k_{ai}^2 - k_{ai}'^2}} \cos \theta' + \sin \theta' \right) \exp \left( -\frac{\hbar^2 k_i k_{ai}}{m_e k_B T_a} \cos \theta' \right)
\]

(2.114)

The angular integral becomes

\[
\int d \theta' \left( \frac{k_{ai}'}{\sqrt{k_{ai}^2 - k_{ai}'^2}} \cos \theta' + \sin \theta' \right) \exp \left( -\frac{\hbar^2 k_i k_{ai}}{m_e k_B T_a} \cos \theta' \right)
= \frac{k_{ai}'}{\sqrt{k_{ai}^2 - k_{ai}'^2}} \int d \theta' \cos \theta' \exp \left( -\frac{\hbar^2 k_i k_{ai}}{m_e k_B T_a} \cos \theta' \right)
= -\frac{2\pi k_{ai}'}{\sqrt{k_{ai}^2 - k_{ai}'^2}} I_1 \left( \frac{\hbar^2 k_i k_{ai}}{m_e k_B T_a} \right)
\]

(2.115)
Therefore

\[ R_{hj2} (i, a \rightarrow j, b) = - n_a \left( \frac{\hbar}{2\pi m_e k_B T_a} \right) \int q dq k_i |W_{ijab} (q)|^2 \]

\[ \int_{k_{ai}'} \frac{k_{ai}' dk_{ai}'}{\sqrt{k_{ai}^2 - k_{ai}'^2}} \exp \left( - \frac{\hbar^2}{2m_e k_B T_a} \left( k_i^2 + k_{ai}'^2 \right) \right) I_1 \left( \frac{\hbar^2 k_i k_{ai}'}{m_e k_B T_a} \right) \]

\[ = - n_a \left( \frac{\hbar}{2\pi m_e k_B T_a} \right) \int q dq k_i |W_{ijab} (q)|^2 \int_0^\infty \frac{k_{ai}' dk}{\sqrt{k^2 + k_{ai}^2}} \exp \left( - \frac{\hbar^2}{2m_e k_B T_a} \left( k_i^2 + k^2 + k_{ai}'^2 \right) \right) I_1 \left( \frac{\hbar^2 k_i \sqrt{k^2 + k_{ai}'^2}}{m_e k_B T_a} \right) \quad (2.116) \]

where \( \kappa^2 = k_{ai}^2 - k_{ai}'^2 \). Next, a sum over the initial distribution is taken to yield

\[ R_{hj2} (i, a \rightarrow j, b) = - n_a \left( \frac{\hbar}{2\pi m_e k_B T_a} \right) \int q dq |W_{ijab} (q)|^2 \int_0^\infty \frac{k_{ai}' dk}{\sqrt{k^2 + k_{ai}^2}} \exp \left( - \frac{\hbar^2}{2m_e k_B T_a} \left( k_i^2 + k^2 + k_{ai}'^2 \right) \right) I_1 \left( \frac{\hbar^2 k_i \sqrt{k^2 + k_{ai}'^2}}{m_e k_B T_a} \right) \]

\[ = - \frac{\hbar^3}{2\pi m_e^2} \left( \frac{n_a}{k_B T_a} \right) \left( \frac{n_i}{k_B T_i} \right) \int q dq |W_{ijab} (q)|^2 \int_0^\infty \frac{k_{ai}' dk_{ai}'}{\sqrt{k^2 + k_{ai}^2}} \exp \left( - \frac{\hbar^2}{2m_e k_B T_a} \left( \kappa^2 + k_{ai}'^2 \right) \right) \]

\[ \int k_i^2 dk_i \exp \left( - \frac{\hbar^2}{2m_e} \left( \frac{1}{k_B T_a} + \frac{1}{k_B T_i} \right) k_i^2 \right) I_1 \left( \frac{\hbar^2 k_i \sqrt{\kappa^2 + k_{ai}'^2}}{m_e k_B T_a} \right) \quad (2.117) \]

The analytic integral

\[ \int_0^\infty x^2 dx \exp (-ax^2) I_1 (bx) \]

\[ = \frac{1}{ab} \left( \frac{b^2}{4a} \right) \exp \left( \frac{b^2}{4a} \right) \quad (2.118) \]
may be used to get

\[ R_{hj2} (i, a \rightarrow j, b) \]

\[ = - \frac{\hbar}{2\pi m_e} \left( \frac{n_a n_i}{k_B (T_a + T_i)} \right) \left( \frac{T_i}{T_a + T_i} \right) \int dq |W_{ijab} (q)|^2 q k'_{ai} \int_0^\infty d\kappa \exp \left( -\frac{\hbar^2 (\kappa^2 + k'^2_{ai})}{2m_e k_B (T_a + T_i)} \right) \]  

(2.119)

Like with the transfer and cooling rates, the integral in \( \kappa \) can be completed analytically, to yield the final answer of

\[ R_{hj2} (i \rightarrow j) \]

\[ = - n_a n_i \frac{1}{2\hbar^2} \sqrt{\frac{m_e}{2\pi k_B (T_a + T_i)}} \left( \frac{T_i}{T_a + T_i} \right) \int dq |W_{ijab} (q)|^2 \frac{\hbar^2 q k'_{ai}}{m_e} \exp \left( -\frac{\hbar^2 k'^2_{ai}}{2m_e k_B (T_a + T_i)} \right) \]  

(2.120)

The total heating rate is given by

\[ R_{hj} (i, a \rightarrow j, b) = R_{ci} (i, a \rightarrow j, b) + R_{hj1} (i, a \rightarrow j, b) + R_{hj2} (i, a \rightarrow j, b) \]  

(2.121)

Exchanging the indices \( i \) and \( a, j \) and \( b \) yields \( R_{hb} \), the heating of the partner’s final subband.

### 2.5 Interactions between bound subbands and continuum states

QCL transport occurs mainly between the bound states that remain confined within the quantum wells. However, there is evidence to suggest that above barrier states play an important role as well. In particular, leakage of bound state electrons to the superlattice continuum is postulated to be a potential cause of temperature degradation.
2.5.1 Distinction between bound and continuum states

At flatband or at low biases, superlattices admit a description in terms of superlattice Bloch states, analogous to those of a bulk solid. Continuum states would be defined as the bands lying above the conduction band barriers. However, QCL transport occurs at much higher fields, where a description in terms of localized Wannier-Stark states is more accurate (the electric field is included in the diagonalization). \[50\]

Bound Wannier-Stark states nested deep within barriers are well estimated by solving Schrodinger’s equation in a truncated domain (essentially using the tight-binding approximation). Unfortunately, attempts to calculate the wavefunctions and energies in this fashion results in complications. A truncated domain (Dirichlet boundary conditions) corresponds to enclosing one or more QCL modules within an infinite potential barrier. This infinite potential barrier possesses its own resonances, which interferes with the spectrum of the superlattice itself, particularly if the computation domain is made too big.

The problem is more than numerical, as there exists fundamental physical differences between bound (below barrier) and continuum states (above barrier). Below barrier states are essentially stationary, although in the presence of an electric field they may develop a small leakage rate due to the probability of tunneling out of the QCL period; this coherent escape process is normally negligible compared to scattering processes and is ignored. In contrast, above-barrier states are never stationary. In the absence of scattering, above barrier electrons will accelerate indefinitely; there is little physical justification in using hard boundary conditions to treat the continuum under bias.

More work is needed in the description of the continuum states in the superlattice. It may necessitate abandoning perturbative transport models and moving to a real-space NEGF based approach or another basis independent quantum method. In the absence of such sophistication, however, the model of Chuan is adopted, in which the continuum states are modelled as plane-waves. \[51\] This assumption admits analytically tractable expressions for scattering rates, and the hope is that integrating...
over the space of all plane-waves will be yield similar results compared to integrating over the detailed spectrum.

2.5.2 Mathematical description of continuum states

A note on notation in the following: because most of this thesis deals with 2D sub-bands, lower case letters will be reserved for 2D momentum quantities (for example \( \mathbf{q} = q_x \hat{x} + q_y \hat{y} \)). Momenta in three 3D will be denoted by upper case letters (for example \( \mathbf{Q} = \mathbf{q} + q_z \hat{z} \)). Like how the subbands are denoted by a lower case letter such as \( a \), the continuum will be taken as a electron reservoir denoted by \( C \).

Using box normalization, the continuum wavefunctions are assumed to have the form

\[
\psi (\mathbf{r}) = \sum_{\beta} e^{i\mathbf{K} \cdot \mathbf{r}} u_{\beta} (\mathbf{r})
\]

where the summation is over the bands, \( \beta = cb, lh, so \). Assuming a QCL with a uniform barrier height (conduction band offset) of \( \Delta E_c \), the energies are assumed to be given by

\[
E (K) = \Delta E_c + \frac{\hbar^2 K^2}{2m_e}
\]

In doing so, all spectral structure in the continuum is neglected. This is a drastic assumption, but it is made for analytic convenience.

Using \( \Delta E_c \) as the energy minimum is only sensible at flatband. Upon application of an electrical bias, \( \Delta E_c \) becomes \( z \)-dependent, and hence ambiguous. As shall be seen in sections 2.6 and 2.7 however, the scattering rates derived at flatband do not depend on \( \Delta E_C \) directly, but on the quantity \( \Delta E_C - E_a \), where \( E_a \) is the subband edge energy of subband \( a \). As a phenomenological extension of the flatband expressions to non-zero bias, the concept of a local barrier height, \( B(z) \) is defined. The local barrier height is illustrated in figure 2.8. The term \( \Delta E_C - E_a \) in flatband scattering expressions is replaced by the wavefunction weighted height

\[
\Delta E_C - E_a \to \int dz (B(z) - E_a) \sum_{\beta} |\phi_{\beta a} (z)|^2
\]
Figure 2-8: Illustration of the “local barrier height” defined for estimating energy distance to continuum when using flatband scattering expression for the biased QCL.

where the summation $\beta$ is over the bands.

The continuum states are assumed to obey a displaced Maxwellian distribution. That is,

$$f_c(K) = \frac{N_c}{2} \left( \frac{2\pi \hbar^2}{m_e k_B T_c} \right)^{3/2} \exp \left( -\frac{\hbar^2 (K - \overline{K}_C)^2}{2m_e k_B T_c} \right)$$

(2.125)

where $\overline{K}_C$ is electric field dependent and reflects the drift velocity of electrons in the continuum.

### 2.6 Single-body scattering between bound subbands and the continuum

This section details the calculation of single-body scattering of bound subband electrons to and from the continuum. Only the LO phonon interaction is covered (see figure 2-9), as it is believed to be the dominant scatterer. Extensions to other scatter-
Figure 2-9: LO phonon interactions with the continuum. (a) Carrier capture through LO phonon emission. (b) Carrier escape through LO phonon absorption.

tering mechanisms should be straightforward, but this is left to future work.

2.6.1 Continuum-to-bound capture through LO phonon emission

Assuming bulk, 3D phonons and interaction through the Frohlich Hamiltonian, the scattering matrix element describing scattering for an electron from the a continuum state $C\mathbf{K}_a$ to a bound state $b\mathbf{K}_b$ by a for a single LO phonon mode is given by

$$
\left| \langle b\mathbf{K}_b | H' | C\mathbf{K}_a \rangle \right|^2 = \frac{1}{V} \frac{e^2 \hbar \omega_{LO}}{2Q^2} \left( \frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_s} \right) \left( N_{LO}(Q) + \frac{1}{2} \pm \frac{1}{2} \right) $$

$$
\frac{1}{L} \delta_{k_a + \mathbf{Q} - k_b} |\Phi_b(k_{za} \mp q_z)|^2
$$

where

$$
\Phi_b(k) = \int dz \left( \sum_\beta \phi_{\beta b}(z) \right) e^{ikz}
$$

is essentially the Fourier transform of the bound subband wavefunction, and the other symbols follow the notation of intersubband LO phonon scattering discussed in section 2.3.5. However, it was found later that the $Q$ in the denominator occasionally
leads to a divergent integral later. Therefore, following ref. [52], the matrix element is amended to include a small amount of static screening (screening details are given in section 3.4.3).

\[
\begin{align*}
\langle b_{\bar{k}_b} | H' | K_{a} \rangle^2 &= \frac{1}{V} \frac{e^2 \hbar \omega_{\text{LO}}}{2} \left( \frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) \left( N_{LO} (Q) + \frac{1}{2} \pm \frac{1}{2} \right) \frac{1}{L} \delta_{k_a + \overline{Q} - K_b} | \Phi_b (k_{za} \mp q_z) |^2 \\
\end{align*}
\]

(2.128)

The total scattering rate summed over all phonon modes is hence

\[
\begin{align*}
W (C K_a \rightarrow b \bar{k}_b) &= \frac{2 \pi}{\hbar} \frac{e^2 \hbar \omega_{\text{LO}}}{2} \left( \frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) \frac{1}{V} \sum_{Q} \frac{Q^2}{(Q^2 + q^2_s)^2} \left( N_{LO} (Q) + \frac{1}{2} \pm \frac{1}{2} \right) \frac{1}{L} \delta_{k_a + \overline{Q} - K_b} | \Phi_b (k_{za} \mp q_z) |^2 \\
&\delta \left( \frac{\hbar^2 K_{za}^2}{2 m_e} - \frac{\hbar^2 \overline{Q}^2}{2 m_e} + \Delta E_c - E_b \mp \hbar \omega_{\text{LO}} \right) \\
\end{align*}
\]

(2.129)

Next, a sum over the initial distribution is taken, and the Kronecker \( \delta \)-function is used to eliminate the integrals in \( k_{za}, k_{ya} \). This leaves an integral in \( k_{za} \). For convenience, a change of variables from \( k_{za} \) to \( k_{zb} = k_{za} \mp q_z \) is applied. Furthermore, a pseudomomentum \( \overline{K}_b = \overline{K}_b + k_{zb} \hat{z} \) may be defined for the final subband, such that

\[
\begin{align*}
W (C \rightarrow b \overline{K}_b) &= \frac{1}{\hbar} \frac{e^2 \hbar \omega_{\text{LO}}}{2} \left( \frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) \frac{N_c}{2} \left( \frac{2 \pi \hbar^2}{m_e k_B T_C} \right)^{3/2} \frac{1}{(2\pi)^3} \\
&\int d^3 \overline{Q} \frac{Q^2}{(Q^2 + q^2_s)^2} \left( N_{LO} (Q) + \frac{1}{2} \pm \frac{1}{2} \right) \left( \overline{K}_b \pm \overline{Q} - K_{C} \right)^2 | \Phi_b (k_{zb}) |^2 \\
&\delta \left( \frac{\hbar^2 (K_{zb} \pm \overline{Q})}{2 m_e} - \frac{\hbar^2 \overline{K}_b^2}{2 m_e} + \Delta E_c - E_b \mp \hbar \omega_{LO} \right) \\
\end{align*}
\]

(2.130)
The term \( \frac{\hbar^2}{2m_e} (\overline{K}_b \pm \overline{Q} - \overline{K}_C)^2 \) can be expanded as

\[
\frac{\hbar^2}{2m_e} (\overline{K}_b \pm \overline{Q} - \overline{K}_C)^2 = \frac{\hbar^2}{2m_e} (\overline{K}_b \pm \overline{Q})^2 - \frac{\hbar^2}{m_e} (\overline{K}_b \pm \overline{Q}) \cdot \overline{K}_C + \frac{\hbar^2 K_C^2}{2m_e}
\]  

(2.131)

Integration against the energy-conserving delta function enables the substitution

\[
\frac{\hbar^2}{2m_e} (\overline{K}_b \pm \overline{Q})^2 \rightarrow \frac{\hbar^2 k_b^2}{2m_e} - \Delta E_c + E_b \pm \hbar \omega_{LO}
\]  

(2.132)

Therefore,

\[
W (C \rightarrow b\overline{k}_b)
= \frac{1}{\hbar} e^2 \hbar \omega_{LO} \left( \frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_s} \right) \frac{N_c}{2} \left( \frac{2\pi \hbar^2}{m_e k_B T_C} \right)^{3/2} \frac{1}{(2\pi)^3} \int dk_{zb} |\Phi_b (k_{zb})|^2 
\exp \left( -\frac{\hbar^2 k_b^2}{2m_e k_B T_C} + \Delta E_c - E_b \mp \hbar \omega_{LO} \right)
\left( \frac{\hbar^2 K_b^2}{k_B T_C} + \frac{\hbar^2 \overline{K}_b \cdot \overline{K}_C}{m_e k_B T_C} \right) \right)
\int d^3 \overline{Q} \frac{Q^2}{(Q^2 + q_s^2)^2} \left( N_{LO} (\overline{Q}) + \frac{1}{2} \pm \frac{1}{2} \right) \exp \left( \mp \frac{\hbar^2 \overline{Q} \cdot \overline{K}_C}{m_e k_B T_C} \right) 
\delta \left( \frac{\hbar^2}{2m_e} (Q^2 \pm 2 \overline{K}_b \cdot \overline{Q}) + \Delta E_c - E_b + \frac{\hbar^2 k_b^2}{2m_e} \pm \hbar \omega_{LO} \right)
\]  

(2.133)

To proceed further, thermal \( N_{LO} (\overline{Q}) = N_{LO} \) is assumed. The integral in \( \overline{Q} \) is completed in spherical coordinates. A coordinate system oriented along \( \overline{K}_b \) is chosen, such that

\[
\overline{K}_b = K_b \hat{k}
\overline{K}_c = K_c \sin \theta \hat{i} + K_c \cos \theta \hat{k}
\overline{Q} = Q \sin \theta \cos \phi \hat{i} + Q \sin \theta \sin \phi \hat{j} + Q \cos \theta \hat{k}
\]  

(2.134)
where $\theta$ is the angle between $\mathbf{Q}$ and $\mathbf{K}_b$, and $\theta'$ is the angle between $\mathbf{K}_b$ and $\mathbf{K}_c$. Therefore,

$$W \left( C \rightarrow bK_b \right) = \frac{N_c}{2} e^2 \hbar \omega_{LO} \left( \frac{1}{\epsilon_\infty} - 1 \right) \left( \frac{2\pi \hbar^2}{m_e k_B T_c} \right)^{3/2} \int dk_z b |\Phi_b (k_z b)|^2$$

$$\exp \left( -\frac{\hbar^2 k_b^2}{2m_e k_B T_c} \frac{\Delta E_c - E_b \mp \hbar \omega_{LO}}{k_B T_c} - \frac{\hbar^2 K_c^2}{2m_e k_B T_c} + \frac{\hbar^2 K_b \cdot \mathbf{K}_c}{m_e k_B T_c} \right)$$

$$\left( N_{LO} + \frac{1}{2} \pm \frac{1}{2} \right) \int \frac{Q^4 dQ}{(Q^2 + q_s^2)^2} \int \sin \theta d\theta d\phi \exp \left( \pm \frac{\hbar^2 Q K_c}{m_e k_B T_c} (\sin \theta \sin \theta' \cos \phi + \cos \theta \cos \theta') \right)$$

$$\delta \left( \frac{\hbar^2}{2m_e} (Q^2 \pm 2K_b Q \cos \theta) + \Delta E_c - E_b + \frac{\hbar^2 k_{zb}^2}{2m_e} \mp \hbar \omega_{LO} \right)$$

The angular integral is completed as

$$\int \sin \theta d\theta d\phi \exp \left( \pm \frac{\hbar^2 Q K_c}{m_e k_B T_c} (\sin \theta \sin \theta' \cos \phi + \cos \theta \cos \theta') \right)$$

$$= \frac{2\pi m_e}{\hbar^2 Q K_b} J_0 \left( \frac{\hbar^2}{2m_e k_B T_c} \frac{K_c}{K_b} \right) \sqrt{4Q^2 K_b^2 - (Q^2 + k^2_{\Delta} + k_{zb}^2 \mp k_{LO}^2)^2} \sin \theta'$$

$$\exp \left( -\frac{\hbar^2}{2m_e k_B T_c} \frac{K_c}{K_b} \right) (Q^2 + k^2_{\Delta} + k_{zb}^2 \mp k_{LO}^2) \cos \theta'$$

$$\Theta \left( |Q^2 + k^2_{\Delta} + k_{zb}^2 \mp k_{LO}^2| < 2Q K_b \right)$$

where $k^2_{\Delta} = \frac{2m_e}{\hbar^2} (\Delta E_c - E_b), k_{LO} = \frac{2m_e}{\hbar^2} (\hbar \omega_{LO})$. The definition of $\theta'$ and the fact that $\mathbf{K}_c = K_c \hat{z}$ implies $\cos \theta' = k_{zb}/K_b$ and $\sin \theta' = k_b/K_b$. Therefore, the becomes
scattering rate

\[ W(C \rightarrow b \overline{F}_b) \]

\[ = N_c \frac{1}{2} \frac{e^2}{\hbar 4\pi \epsilon_0} \left( \frac{\epsilon_0 - \epsilon_s}{\epsilon_\infty - \epsilon_s} \right) \left( \frac{2\pi \hbar^2}{m_e k_B T_c} \right)^{1/2} \left( \frac{\hbar \omega_{LO}}{k_B T_c} \right) \int dk_{zb} |\Phi_b(k_{zb})|^2 \]

\[ \frac{1}{K_b} \exp \left( -\frac{\hbar^2}{2m_e k_B T_c} \left( k_b^2 - k_b^2 \pm k_{LO}^2 + K_c^2 - 2K_c k_{zb} \right) \right) \]

\[ \int \frac{Q^3 dQ}{(Q^2 + q_s^2)^2} \left( N_{LO} \left( \frac{Q}{Q_{LO}} + 1 \right) + 1 \right) \]

\[ I_0 \left( \frac{\hbar^2}{2m_e k_B T_c} \left( \frac{K_c k_{zb}}{K_b^2} \right) \sqrt{4Q^2 K_b^2 - (Q^2 + k_{LO}^2 + k_{zb}^2 + k_{LO}^2)} \right) \]

\[ \exp \left( -\frac{\hbar^2}{2m_e k_B T_c} \left( \frac{K_c k_{zb}}{K_b^2} \right) \left( Q^2 + k_{LO}^2 + k_{zb}^2 + k_{LO}^2 \right) \right) \]

\[ \Theta \left( |Q^2 + k_{LO}^2 + k_{zb}^2| < 2QK_b \right) \]

Up to this point, the results are exact, but I have found no further means to continue analytically. Furthermore, numerically, the inclusion of the \( K_c \) terms creates divergent integrals, for reasons not well understood. Going forward, the \( K_c \) terms are ignored. This yields

\[ W(C \rightarrow b \overline{F}_b) \]

\[ = N_c \frac{1}{2} \frac{e^2}{\hbar 4\pi \epsilon_0} \left( \frac{\epsilon_0 - \epsilon_s}{\epsilon_\infty - \epsilon_s} \right) \left( \frac{2\pi \hbar^2}{m_e k_B T_c} \right)^{1/2} \left( \frac{\hbar \omega_{LO}}{k_B T_c} \right) \int dk_{zb} |\Phi_b(k_{zb})|^2 \]

\[ \frac{1}{K_b} \exp \left( -\frac{\hbar^2}{2m_e k_B T_c} \left( k_b^2 - k_b^2 \pm k_{LO}^2 \right) \right) \]

\[ \left( N_{LO} + \frac{1}{2} \pm \frac{1}{2} \right) \int \frac{Q^3 dQ}{(Q^2 + q_s^2)^2} \]

\[ \Theta \left( |Q^2 + k_{LO}^2 + k_{zb}^2| < 2QK_b \right) \]

The integral in \( Q \) can now be completed. The step function defines the maximum and minimum \( Q \) in the integration.

\[ Q_{\text{min}} = k_b - \sqrt{k_b^2 - k_{LO}^2 + k_{zb}^2} = k_b + k_b^2 + k_{LO}^2 \]

\[ Q_{\text{max}} = k_b + \sqrt{k_b^2 - k_{LO}^2 + k_{zb}^2} = k_b + k_b^2 + k_{LO}^2 \]

(2.139)
with solution possible only $k_b^2 > k^2_{\Delta} \mp k^2_{LO}$. Therefore,

$$
\int \frac{Q^3 dQ}{(Q^2 + q_s^2)^2} \Theta \left( |Q^2 + k^2_{\Delta} + k^2_{zb} \mp k^2_{LO}| < 2Qk_b \right)
$$

$$
= \Theta \left( k_b^2 > k^2_{\Delta} \mp k^2_{LO} \right) \frac{1}{2} \left[ g(Q_{\text{max}}) - g(Q_{\text{min}}) \right]
$$

(2.140)

where the function $g$ is given by

$$
g(Q) = \frac{q_s^2}{Q^2 + q_s^2} + \ln \left( Q^2 + q_s^2 \right)
$$

(2.141)

Therefore,

$$
W(C \to b)
$$

$$
= \frac{N_c e^2}{4\pi \hbar 4\pi \epsilon_0} \left( \frac{\epsilon_0 - \epsilon_s}{\epsilon_\infty} \right) \left( \frac{2\pi \hbar^2}{me k_B T_c} \right)^{1/2} \left( \frac{\hbar \omega_{LO}}{k_B T_c} \right) \left( N_{LO} + \frac{1}{2} \pm \frac{1}{2} \right)
$$

$$
\int dk \int_0^\infty dk_b \Phi_b(kzb) \left| f \right|^2 \exp \left( -\frac{\hbar^2}{2me k_B T_c} \left( k_b^2 - k^2_{\Delta} \mp k^2_{LO} \right) \right)
$$

$$
\Theta \left( k_b^2 > k^2_{\Delta} \mp k^2_{LO} \right) \left[ g(Q_{\text{max}}) - g(Q_{\text{min}}) \right]
$$

(2.142)

The upper sign is taken for LO emission, and the heating rate is given by insertion of a term $\frac{\hbar^2 k^2_b}{2me}$ into the integrand above. The cooling rate of the continuum will be neglected, for reasons expressed in section 3.6.1.
2.6.2 Bound-to-continuum escape through LO phonon absorption

The derivation is similar to equation 2.142. The transfer rate is given by

\[ W(a \rightarrow C) = \frac{n_a e^2}{4\pi\hbar 4\pi\varepsilon_0} \left( \frac{\varepsilon_0 - \varepsilon_s}{\varepsilon_\infty} \right) \left( \frac{\hbar \omega_{\text{LO}}}{k_B T_a} \right) \left( N_{\text{LO}} + \frac{1}{2} \pm \frac{1}{2} \right) \int \! dk_{za} \int \! dz \psi_a(z) e^{ik_{za}z} \left( \int \! dk_a \frac{k_a}{2m_e k_B T_a} \exp \left( -\frac{\hbar^2 k_a^2}{2m_e k_B T_a} \right) \right) \Theta (k_a^2 > k_{\Delta}^2 + k_{\text{LO}}^2) \left[ g (Q_{\text{max}}) - g (Q_{\text{min}}) \right] \]

(2.143)

where

\[ Q_{\text{min}} = K_a - \sqrt{K_a^2 - k_{\Delta}^2 - k_{za}^2 \pm k_{\text{LO}}^2} \]

(2.144)

\[ Q_{\text{max}} = K_a + \sqrt{K_a^2 - k_{\Delta}^2 - k_{za}^2 \pm k_{\text{LO}}^2} \]

(2.145)

and the function \( g \) is given by

\[ g (Q) = \frac{q_s^2}{Q^2 + q_s^2} + \ln (Q^2 + q_s^2) \]

(2.146)

The cooling rate is given by the insertion of a term \( \frac{\hbar^2 k_a^2}{2m_e} \) into the integrand above. The heating rate of the continuum will be neglected, as the continuum temperature in this work is modeled differently from the subbands (see section 3.6.1).

2.6.3 Verification of detailed balance

In equilibrium, equations 2.142 and 2.143 should be equal according the principle of detailed balance. As a sanity-check, this is shown here to be the case.

In equilibrium, the continuum and bound subbands must share the same temper-
nature and chemical potential. This implies

\[ T_c = T_a = T \]  \hspace{1cm} (2.147)

\[ n_a = N_C \sqrt{\frac{2 \pi \hbar^2}{m_e k_B T}} e^{(\Delta E_c - E_a)/k_B T} \]  \hspace{1cm} (2.148)

This yields for equation 2.142

\[
W(C \to b) = \frac{N_C e^2}{4 \pi \hbar 4 \pi \epsilon_0} \left( \frac{\epsilon_0 - \epsilon_s}{\epsilon_\infty - \epsilon_s} \right) \left( \frac{2 \pi \hbar^2}{m_e k_B T} \right)^{1/2} \left( \frac{\hbar \omega_{LO}}{k_B T_c} \right) (N_{LO} + 1) \exp \left( \frac{\Delta E_c - E_a - \hbar \omega_{LO}}{k_B T_c} \right) \\
\int dk_{zb} |\Phi_b(k_{zb})|^2 \int_0^\infty dk_b k_b K_b \exp \left( -\frac{\hbar^2 k_b^2}{2 m_e k_B T} \right) \\
\Theta \left(k_b^2 > k_\Delta^2 \mp k_{LO}^2\right) [g(Q_{max}) - g(Q_{min})]
\]  \hspace{1cm} (2.149)

and for equation 2.143

\[
W(a \to C) = \frac{N_C e^2}{4 \pi \hbar 4 \pi \epsilon_0} \left( \frac{\epsilon_0 - \epsilon_s}{\epsilon_\infty - \epsilon_s} \right) \left( \frac{2 \pi \hbar^2}{m_e k_B T} \right)^{1/2} \left( \frac{\hbar \omega_{LO}}{k_B T_a} \right) N_{LO} e^{(\Delta E_c - E_a)/k_B T_c} \\
\int dk_{za} \left| \int dz \psi_a(z) e^{ik_{za} z} \right|^2 \int dk_a k_a K_a \exp \left( -\frac{\hbar^2 k_a^2}{2 m_e k_B T} \right) \\
\Theta \left(k_a^2 > k_\Delta^2 \mp k_{LO}^2\right) [g(Q_{max}) - g(Q_{min})]
\]  \hspace{1cm} (2.150)

For temperatures much less than the LO phonon energy, \( N_{LO} \approx \exp \left( -\frac{\hbar \omega_{LO}}{k_B T} \right) \) and \( N_{LO} + 1 \approx 1 \), so the two expressions are equal, as expected.

### 2.7 Two-body interactions between bound subbands and the continuum

Similar to the intersubband case, electrons can again scatter off other electrons when interacting with continuum states. This section concentrates in particular on inter-
Figure 2-10: Electron-electron scattering between continuum and bound electrons. (a) Intersubband impact ionization: continuum electron liberates bound electron during collision. (b) Intersubband Auger recombination: two continuum electrons collide, one becomes bound due to energy loss to the other.

*subband impact ionization* (ISII), in which electrons in bound subband are knocked loose by a high energy continuum electron. The inverse process of *intersubband Auger recombination* (ISAR) is also studied for physical symmetry. These processes are illustrated in figure 2-10.

Impact ionization in general is the process by which a high energy free electron collides with and liberates a free electron. This process is well studied for avalanche photomultiplication in bulk semiconductors. In a QCL or superlattice, the bound subbands may be regarded as a “valence band” of sorts, and continuum states as a “conduction band.”
2.7.1 Intersubband impact ionization

Following ref. [51], Fermi’s golden rule is used for our analysis. The quantity of interest is

\[ W(C, a \rightarrow C, C') = 2 \sum_{K_i} \sum_{k_a} \sum_{K_j} \sum_{K_b} W(CK_i, aK_a \rightarrow C K_j, C K_b) f_C(K_i) f_a(k_a) (1 - f_C(K_j)) (1 - f_C(K_b)) \]

\[ \approx 2 \sum_{K_i} \sum_{k_a} \sum_{K_j} \sum_{K_b} W(CK_i, aK_a \rightarrow C K_j, C K_b) f_C(K_i) f_a(k_a) \]

(2.151)

where state-blocking has been ignored. The above equation describes an 11-dimensional integral. The presence of δ-functions in the scattering rate \( W(CK_i, aK_a \rightarrow C K_j, C K_b) \) (see below) reduces this number to 8 non-trivial integration dimensions, but this is still far too high a number for efficient numerical calculation. Similar to section 2.4, Boltzmann statistics are used to reduce the integral from 8-dimensions to 2.

For a simple, isotropically screened coulomb potential, the matrix element for scattering is given by

\[ \left| \langle CK_i, aK_a \rangle \frac{-e^2}{4\pi \epsilon r} \exp(-qs r) | CK_j C K_b \rangle \right|^2 \]

\[ = \frac{1}{L} \left( \frac{e^2}{eV} \right)^2 \frac{|\Phi_a(k_{zb} - k_{zi} + k_{zj})|^2}{\left[(K_i - K_j)^2 + q_s^2\right]^2} \delta_{\bar{K}_j = \bar{K}_i - \bar{q}} \]

(2.152)

Fermi’s golden rule for impact ionization is therefore given by

\[ W(CK_i, aK_a \rightarrow C K_j, C K_b) \]

\[ = \frac{2\pi}{\hbar} \frac{1}{L} \left( \frac{e^2}{eV} \right)^2 \frac{|\Phi_a(k_{zb} - k_{zi} + k_{zj})|^2}{\left[(K_i - K_j)^2 + q_s^2\right]^2} \delta_{\bar{K}_j = \bar{K}_i - \bar{q}} \]

\[ \delta \left( \frac{\hbar^2 K_i^2}{2m_e} + \frac{\hbar^2 k_a^2}{2m_e} - \frac{\hbar^2 K_j^2}{2m_e} - \frac{\hbar^2 K_b^2}{2m_e} - (\Delta E_c - E_a) \right) \]

(2.153)

Next, a sum over final states \( \bar{K}_j \) is taken. The δ-function is used to remove the
integrals in $k_{xj}$ and $k_{yj}$. For the remaining integral $k_{zj}$, a change of coordinates to $k_{za} = k_{zb} - k_{zi} + k_{zj}$ is made, and $q_z = k_{zi} - k_{xj}$ is defined. This yields

$$W (C \overline{K}_i, a \overline{K}_a \rightarrow C, C \overline{K}_b)$$

$$= \frac{1}{\hbar} \left( \frac{e^2}{\epsilon V} \right)^2 \frac{1}{V} \int dk_{za} \frac{1}{(2\pi)^3} \int \frac{Q^2}{Q^2 + q_z^2} \left[ \Phi_a (k_{za}) \right]^2$$

$$\delta \left( \frac{\hbar^2}{2m_e} (Q^2 - \overline{Q} \cdot (\overline{K}_i - \overline{K}_a)) + \left( \Delta E_c - E_a + \frac{\hbar^2 k_{za}^2}{2m_e} \right) \right)$$

(2.154)

where the pseudo-momentum $\overline{K}_a$ is given by $\overline{K}_a = \overline{k}_a + k_{za} \hat{z}$, and pseudo-exchanged momentum by $\overline{Q} = \overline{K}_b - \overline{K}_a$. Next, a sum over the final states of the partner is taken, through a summation in $\overline{Q}$. A coordinate system aligned along $\overline{K}_{ia} = \overline{K}_i - \overline{K}_a$ is chosen for integration in spherical coordinates, such that $\overline{Q} \cdot (\overline{K}_i - \overline{K}_a) = QK_{ia} \cos \theta$.

Then

$$W (C \overline{K}_i, a \overline{K}_a \rightarrow C, C)$$

$$= \frac{1}{\hbar} \left( \frac{e^2}{\epsilon V} \right)^2 \frac{1}{V} \int dk_{za} \frac{1}{(2\pi)^3} \int \frac{Q^2}{Q^2 + q_z^2} \left[ \Phi_a (k_{za}) \right]^2$$

$$\delta \left( \frac{\hbar^2}{m_e} (Q^2 - QK_{ia} \cos \theta) + \left( \Delta E_c - E_a + \frac{\hbar^2 k_{za}^2}{2m_e} \right) \right)$$

(2.155)

The angular integrals may be completed by defining $u = \cos \theta$, to yield

$$\int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi \delta \left( \frac{\hbar^2}{m_e} (Q^2 - QK_{ia} \cos \theta) + \left( \Delta E_c - E_a + \frac{\hbar^2 k_{za}^2}{2m_e} \right) \right)$$

$$= \frac{2\pi m_e}{\hbar^2 QK_{ia}} \int_{-1}^1 du \delta \left( \frac{Q}{K_{ia}} - u + \frac{m_e}{\hbar^2 QK_{ia}} \left( \Delta E_c - E_a + \frac{\hbar^2 k_{za}^2}{2m_e} \right) \right)$$

$$= \frac{2\pi m_e}{\hbar^2 QK_{ia}} \Theta \left( \left\{ \frac{Q}{K_{ia}} + \frac{m_e}{\hbar^2 QK_{ia}} \left( \Delta E_c - E_a + \frac{\hbar^2 k_{za}^2}{2m_e} \right) \right\} < 1 \right)$$

$$= \frac{2\pi m_e}{\hbar^2 QK_{ia}} \Theta \left( \left| Q^2 + \frac{1}{2} (k_{\Delta}^2 + k_{za}^2) \right| < QK_{ia} \right)$$

(2.156)
where \( k_\Delta^2 = \frac{2m_e}{\hbar^2} (\Delta E_c - E_a) \). The argument of the \( \Theta \) function defines a maximum and minimum \( Q \) of

\[
Q_{\text{max}} = \frac{1}{2} \left[ K_{ia} + \sqrt{K_{ia}^2 - 2 (k_\Delta^2 + k_{za}^2)} \right] \\
Q_{\text{min}} = \frac{1}{2} \left[ K_{ia} - \sqrt{K_{ia}^2 - 2 (k_\Delta^2 + k_{za}^2)} \right]
\]

(2.157)

(2.158)

Furthermore, a solution is only possible if

\[
K_{ia}^2 > 2 (k_\Delta^2 + k_{za}^2)
\]

(2.159)

Inserting these results into the scattering rate, and completing the integral in \( Q \), the rate becomes

\[
W (C\overline{K}_i, a\overline{k}_a \rightarrow C, C') \\\n= \frac{1}{\hbar} \left( \frac{e^2}{4\pi \epsilon} \right)^2 \frac{2m_e}{\hbar^2} \frac{1}{V} \int dk_{za} |\Phi_a (k_{za})|^2 \]

\[
\frac{1}{K_{ia}} \Theta (K_{ia}^2 > 2 (k_\Delta^2 + k_{za}^2)) \left[ \frac{1}{Q_{\text{min}}^2 + q_s^2} - \frac{1}{Q_{\text{max}}^2 + q_s^2} \right]
\]

(2.160)

Next, a sum over the initial distribution of the continuum electrons is taken through a sum in \( \overline{K}_{ia} \). Using the variable \( \overline{K}_{ia} \), the continuum distribution function is

\[
f_C (\overline{K}_i) = \frac{N_C}{2} \left( \frac{2\pi \hbar^2}{m_e k_B T_C} \right)^{3/2} \exp \left( -\frac{\hbar^2}{2m_e k_B T_C} (\overline{K}_i - \overline{K}_c)^2 \right) \]

\[
= \frac{N_C}{2} \left( \frac{2\pi \hbar^2}{m_e k_B T_C} \right)^{3/2} \exp \left( -\frac{\hbar^2}{2m_e k_B T_C} (\overline{K}_{ia} + \overline{K}_{ac})^2 \right) \]

\[
= \frac{N_C}{2} \left( \frac{2\pi \hbar^2}{m_e k_B T_C} \right)^{3/2} \exp \left( -\frac{\hbar^2}{2m_e k_B T_C} (K_{ia}^2 + K_{ca}^2 + 2K_{ia}K_{ac} \cos \theta') \right)
\]

(2.161)

where \( \theta' \) is the angle between \( \overline{K}_{ia} \) and \( \overline{K}_{ac} = \overline{K}_a - \overline{K}_c \). Thus, the scattering rate
becomes

\[
W(C, a \vec{k}_a \to C, C) = \frac{N_C}{2 \pi \hbar} \left( \frac{e^2}{4 \pi \epsilon} \right)^2 \left( \frac{2 \pi \hbar^2}{me k_B T_C} \right)^{1/2} \left( \frac{1}{k_B T_C} \right) \int dk_{za} |\Phi_a(k_{za})|^2 \int_{\sqrt{2(k_{za}^2 + k_{za}^2)}}^\infty K_{ia} dK_{ia} \\
\int \sin \theta' d\theta' \exp \left( -\frac{h^2}{2me k_B T_C} \left( K_{ia}^2 + K_{ca}^2 + 2K_{ia}K_{ac} \cos \theta' \right) \right) \\
\left[ \frac{1}{Q_{\text{min}}^2 + q_s^2} - \frac{1}{Q_{\text{max}}^2 + q_s^2} \right]
\]

(2.162)

The angular integral can be completed using the substitution \( v = \sin \theta' \), to yield

\[
\int \sin \theta' d\theta' \exp \left( -\frac{h^2}{m e k_B T_C} \cos \theta' \right) = \int_{-1}^{1} dv \exp \left( -\frac{h^2}{m e k_B T_C} v \right) \\
= \frac{m e k_B T_C}{h^2 K_{ia} K_{ca}} \left[ \exp \left( \frac{h^2 K_{ia} K_{ac}}{m e k_B T_C} \right) - \exp \left( -\frac{h^2 K_{ia} K_{ac}}{m e k_B T_C} \right) \right]
\]

(2.163)

Therefore, the scattering rate becomes

\[
W(C, a \vec{k}_a \to C, C) = \frac{N_C}{\hbar^2} \left( \frac{e^2}{4 \pi \epsilon} \right)^2 \left( \frac{m_e}{2 \pi k_B T_C} \right)^{1/2} \int dk_{za} |\Phi_a(k_{za})|^2 \int_{\sqrt{2(k_{za}^2 + k_{za}^2)}}^\infty \frac{dK_{ia}}{K_{ca}} \\
\exp \left( -\frac{h^2}{2m e k_B T_C} \left( K_{ia}^2 + K_{ca}^2 \right) \right) \\
\left[ \exp \left( \frac{h^2 K_{ia} K_{ac}}{m e k_B T_C} \right) - \exp \left( -\frac{h^2 K_{ia} K_{ac}}{m e k_B T_C} \right) \right] \\
\left[ \frac{1}{Q_{\text{min}}^2 + q_s^2} - \frac{1}{Q_{\text{max}}^2 + q_s^2} \right]
\]

(2.164)
To decouple the integration limits, let $\kappa^2 = K_{ia}^2 - 2(k_\Delta^2 + k_{za}^2)$. Then

$$dK_{ia} = \frac{\kappa d\kappa}{\sqrt{\kappa^2 + 2(k_\Delta^2 + k_{za}^2)}} = \frac{\kappa d\kappa}{K_{ia}}. \quad (2.165)$$

Furthermore, the term involving $Q_{\text{min}}$ and $Q_{\text{max}}$ simplifies to

$$\frac{1}{Q_{\text{min}}^2 + q_s^2} - \frac{1}{Q_{\text{max}}^2 + q_s^2} = \frac{2K_{ia}^2}{2q_s^2\kappa^2 + (k_\Delta^2 + k_{za}^2 + q_s^2)^2} \quad (2.166)$$

Then

$$W(C, a\overline{k}_a \rightarrow C, C) = N_C \frac{2}{\hbar^2} \left( \frac{e^2}{4\pi\epsilon} \right)^2 \left( \frac{m_e}{2\pi k_B T_C} \right)^{1/2} \int dk_{za} |\Phi_a(k_{za})|^2 \int_0^\infty \frac{\kappa^2 d\kappa}{2q_s^2\kappa^2 + (k_\Delta^2 + k_{za}^2 + q_s^2)^2} \exp \left( -\frac{\hbar^2}{2m_e k_B T_C} (K_{ia}^2 + K_{ca}^2) \right) \frac{1}{K_{ca}} \left[ \exp \left( \frac{\hbar^2 K_{ia} K_{ac}}{m_e k_B T_C} \right) - \exp \left( -\frac{\hbar^2 K_{ia} K_{ac}}{m_e k_B T_C} \right) \right] \quad (2.167)$$

Finally, a sum over the distribution of the subband $a$ is taken, dividing by area and accounting for spin. At this point, $K_{ia}^2 = \kappa^2 + 2(k_\Delta^2 + k_{za}^2)$ and $K_{ac}^2 = k_a^2 + (k_{za} - K_c)^2$ are both independent of angle.

$$W(C, a\overline{k}_a \rightarrow C, C) = 2N_C n_a \left( \frac{e^2}{4\pi\epsilon} \right)^2 \left( \frac{1}{2\pi m_e k_B T_C} \right)^{1/2} \left( \frac{1}{k_B T_a} \right) \int dk_{za} |\Phi_a(k_{za})|^2 \int_0^\infty \frac{\kappa^2 d\kappa}{2q_s^2\kappa^2 + (k_\Delta^2 + k_{za}^2 + q_s^2)^2} \int k_a dk_a \exp \left( -\frac{\hbar^2 k_a^2}{2m_e k_B T_a} \right) \exp \left( -\frac{\hbar^2}{2m_e k_B T_C} (K_{ia}^2 + K_{ca}^2) \right) \frac{1}{K_{ca}} \left[ \exp \left( \frac{\hbar^2 K_{ia} K_{ac}}{m_e k_B T_C} \right) - \exp \left( -\frac{\hbar^2 K_{ia} K_{ac}}{m_e k_B T_C} \right) \right] \quad (2.168)$$

84
A change variables is made from \( k_a \) to \( K_{ac}^2 = k_a^2 + (k_{za} - K_C)^2 \).

\[
W(C, a \rightarrow C, C) = 2N_C n_a \left( \frac{e^2}{4\pi\epsilon} \right)^2 \left( \frac{1}{2\pi m_e k_B T_C} \right)^{1/2} \left( \frac{1}{k_B T_a} \right) \int dk_{za} |\phi_a(k_{za})|^2 \int_0^\infty \frac{\kappa^2 d\kappa}{2q^2 \kappa^2 + (k^2 + k_{za}^2 + q_s^2)^2} \exp \left( -\frac{\hbar^2 K_{ia}^2}{2m_e k_B T_C} + \frac{\hbar^2 (k_{za} - K_C)^2}{2m_e k_B T_a} \right) \left[ \exp \left( \frac{\hbar^2 K_{ia} K_{ac}}{m_e k_B T_C} \right) - \exp \left( -\frac{\hbar^2 K_{ia} K_{ac}}{m_e k_B T_C} \right) \right]
\]

(2.169)

To continue, the analytic integral

\[
\int_c^\infty dx e^{-\alpha(x^2 + 2\beta y)} = \frac{1}{2} \sqrt{\frac{\pi}{\alpha}} \exp \left( \alpha \beta^2 y^2 \right) \text{erfc} \left( \sqrt{\alpha} (c + \beta y) \right)
\]

(2.170)

is used to arrive at the final answer

\[
W(C, a \rightarrow C, C) = N_C n_a \frac{1}{\hbar} \left( \frac{e^2}{4\pi\epsilon} \right)^2 \sqrt{\frac{\beta}{k_B T_a}} \int dk_{za} \left| \phi_a(k_{za}) \right|^2 \int_0^\infty \frac{\kappa^2 d\kappa}{2q^2 \kappa^2 + (k^2 + k_{za}^2 + q_s^2)^2} \exp \left( -\alpha \left( 1 - \beta \right) \left( \beta K_{ia}^2 - (k_{za} - K_C)^2 \right) \right) \left[ \text{erfc} \left( \chi_- \right) - \text{erfc} \left( \chi_+ \right) \right]
\]

(2.171)

where

\[
\alpha = \frac{\hbar^2}{2m_e k_B} \left( \frac{1}{T_a} + \frac{1}{T_C} \right)
\]

(2.172)

\[
\beta = \frac{T_a}{T_a + T_C}
\]

(2.173)

\[
\chi_\pm = \sqrt{\alpha} \left( |k_{za} - K_C| \pm \beta K_{ia} \right)
\]

(2.174)
Next, the cooling rate of the ionized subband is considered. This is given by equation 2.168 except that a factor of $\hbar^2 k_a^2 \over 2m_e$ is inserted into the integrand. This yields

$$R_c (C, a \to C, C)$$

$$= 2N_C n_a \left( {e^2 \over 4\pi \epsilon} \right)^2 \left( {1 \over 2\pi m_e k_B T_C} \right)^{1/2} \left( {1 \over k_B T_a} \right) \int dk_{za} |\Phi_a (k_{za})|^2 \int_0^\infty {k^2 dk} \int_{2q_s^2 k^2 + (k_a^2 + k_{za}^2 + q_s^2)^2}^\infty \exp \left( -{\hbar^2 K_{ia}^2 \over 2m_e k_B T_C} + {\hbar^2 |k_{za} - K_c|^2 \over 2m_e k_B T_a} \right)$$

$$\int \exp \left( -{\hbar^2 K_{ia} K_{ac} \over m_e k_B T_C} \right) \left[ \exp \left( -{\hbar^2 K_{ia} K_{ac} \over m_e k_B T_C} \right) - \exp \left( -{\hbar^2 K_{ia} K_{ac} \over m_e k_B T_C} \right) \right]$$

As before, a change of variables is used to integrate in $K_{ac} = \sqrt{k_a^2 + (k_{za} - K_c)^2}$ instead of $k_a$. Therefore, the cooling rate is rewritten as

$$R_c (C, a \to C, C)$$

$$= 2N_C n_a \left( {e^2 \over 4\pi \epsilon} \right)^2 \left( {1 \over 2\pi m_e k_B T_C} \right)^{1/2} \left( {1 \over k_B T_a} \right) \int dk_{za} |\Phi_a (k_{za})|^2 \int_0^\infty {k^2 dk} \int_{2q_s^2 k^2 + (k_a^2 + k_{za}^2 + q_s^2)^2}^\infty \exp \left( -{\hbar^2 K_{ia}^2 \over 2m_e k_B T_C} + {\hbar^2 |k_{za} - K_c|^2 \over 2m_e k_B T_a} \right)$$

$$\int \exp \left( -{\hbar^2 K_{ia} K_{ac} \over m_e k_B T_C} \right) \left[ \exp \left( -{\hbar^2 K_{ia} K_{ac} \over m_e k_B T_C} \right) - \exp \left( -{\hbar^2 K_{ia} K_{ac} \over m_e k_B T_C} \right) \right]$$

To continue, the analytic integral

$$\int_{-\infty}^{\infty} x^2 dx e^{-\alpha(x^2 + 2\beta y)}$$

$$= \frac{1}{2\alpha} \left[ (c + \beta y) \exp \left( -\alpha (c + \beta y)^2 \right) \right]$$

$$+ \frac{1}{2} \sqrt{\frac{\pi}{\alpha}} \left( 1 + 2\alpha \beta^2 y^2 \right) \operatorname{erfc} \left( \sqrt{\alpha} (c + \beta y) \right)$$

(2.177)
can be used to arrive at the final result

\[ R_c (C, a \rightarrow C, C) = N_C n_a \frac{1}{\hbar} \left( \frac{e^2}{4\pi\epsilon} \right)^2 \sqrt{\beta} (1 - \beta) \]

\[
= N_C n_a \frac{1}{\hbar} \left( \frac{e^2}{4\pi\epsilon} \right)^2 \sqrt{\beta} (1 - \beta) \\
\int dk_{za} |\Phi_a (k_{za})|^2 \int_0^\infty \frac{\kappa^2 d\kappa}{2q_s^2 \kappa^2 + (k_{\Delta}^2 + k_{za}^2 + q_s^2)^2} \\
\exp \left( -\alpha (1 - \beta) (\beta K_{ia}^2 - (k_{za} - K_c)^2) \right) \\
\left\{ \frac{1}{\sqrt{\pi}} \left[ \chi_+ \exp \left( -\chi_+^2 \right) - \chi_- \exp \left( -\chi_-^2 \right) \right] \\
+ \left( \frac{1}{2} - \chi_-\chi_+ \right) \left[ \text{erfc} (\chi_-) - \text{erfc} (\chi_+) \right] \right\}
\]

As in the single body case, the heating rate of the continuum will not be explicitly calculated.

### 2.7.2 Intersubband Auger recombination

The inverse process of impact ionization is Auger recombination. In this process, a continuum electron falls into a bound state by relinquishing its energy to another continuum electron. The matrix element for scattering is the same as impact ionization.

The derivation is similar to that for impact ionization, so this section only sketches the main steps and states the final result. The first quantity of interest, the transfer rate, is given by

\[
W (C, C \rightarrow C, b) = \frac{2}{A} \sum \kappa_i \sum \kappa_a \sum \kappa_j \sum \kappa_b W (C \kappa_i, C \kappa_a \rightarrow C \kappa_j, b \kappa_b) \\
\approx \frac{2}{A} \sum \kappa_i \sum \kappa_a \sum \kappa_j \sum \kappa_b W (C \kappa_i, C \kappa_a \rightarrow C \kappa_j, b \kappa_b) f_C (\kappa_i) f_C (\kappa_a)
\]

The derivation proceeds as follows.
1. Start with Fermi’s golden rule.

2. Sum over the initial distribution of the partner electron \( \sum_{K_i} f_C (K_i) \cdot \cdot \cdot \), employing momentum conservation \( \delta \)-function. Use a change of variables to define \( k_{zb} = k_{za} + k_{zi} - k_{zj} \), and the pseudo-momenta \( \overline{K}_b = k_b + k_{zb} \).

3. Sum over the initial distribution of the primary electron, \( \sum_{K_a} f_C (K_a) \cdot \cdot \cdot \). But use a change of variables to complete the integration in the exchanged momentum \( \overline{Q} \) rather than \( K_a \) directly.

4. Sum over the final states of the partner, \( \sum_{K_j} \cdot \cdot \cdot \). But use a change of variables to complete the integration in \( K_{jb} = K_j - \overline{K}_b \) rather than \( K_j \).

5. Sum over the final states, dividing by area and accounting for spin \( \frac{3}{4} \sum_{K_b} \).

The end result is

\[
W (C, C \rightarrow C, b) = N_c^2 \left( \frac{e^2}{4 \pi \epsilon} \right)^2 \left( \frac{1}{k_B T_c} \right)^2 \sqrt{\frac{\pi k_B T_c}{m_e}} \int_{-\infty}^{\infty} dk_{zb} |\Phi_b (k_{zb})|^2 \int_0^\infty \frac{k^2 dk}{2q_s^2 \kappa^2 + (k^2 + k_{zb}^2 + q_s^2)^2} \exp \left( -\frac{1}{4} \alpha \kappa^2 \right) \left[ \text{erfc} (\chi_-) - \text{erfc} (\chi_+) \right] \tag{2.180}
\]

where

\[
k^2_\Delta = \frac{2m_e}{\hbar^2} (\Delta E_c - E_b) \tag{2.181}
\]

\[
\alpha = \frac{k^2}{m_e k_B T_c} \tag{2.182}
\]

\[
K_{jb} = \sqrt{k^2 + 2k^2_\Delta + 2k_{zb}^2} \tag{2.183}
\]

\[
\chi_\pm = \sqrt{\alpha} \left( |k_{zb} - K_c| \pm \frac{1}{2} K_{jb} \right) \tag{2.184}
\]
Similar to equation 2.171, the heating of the destination subband is given by

\[
R_h (C, C \rightarrow C, b) = N_c^2 \left( \frac{e^2}{4 \pi \epsilon} \right)^2 \left( \frac{1}{k_B T_c} \right)^2 \sqrt{\frac{\pi k_B T_c}{m_e}} \left( \frac{1}{2} k_B T_c \right) \int_{-\infty}^{\infty} dk_z b |\Phi_b (k_z b)|^2 \int_0^\infty \frac{\kappa^2 d\kappa}{2 q_s^2 \kappa^2 + (k^2_\Delta + k^2_z + q^2_s)^2} \exp \left( -\frac{1}{4} \alpha \kappa^2 \right) \left\{ \frac{1}{\sqrt{\pi}} \left[ \chi_+ \exp (-\chi_-^2) - \chi_- \exp (-\chi_+^2) \right] + \left( \frac{1}{2} - \chi_- \chi_+ \right) [\text{erfc} (\chi_-) - \text{erfc} (\chi_+)] \right\} \}
\]

(2.185)

2.7.3 Verification of detailed balance

Similar to 2.6.3, equations 2.171 and 2.180 can also be easily verified to satisfy detailed balance in equilibrium. Again, the temperatures and Fermi levels must be the same, but the additional condition \( K_c = 0 \) is required (i.e. there is no net electron motion in the continuum).

2.8 Conclusion

A method for calculating the eigenstates in THz QCLs is given, along with the necessary expressions to describe tunneling and scattering transport between them. By assuming Boltzmann statistics for the superlattice subbands, transport integrals are significantly simplified compared to popular expressions found in the literature, in most cases requiring only a single 1D integral. In particular, a reformulation of inter-subband electron-electron scattering is given that is more computationally efficient than common expressions in the literature.

Furthermore, scattering between bound and continuum states due to LO phonon interactions and electron-electron scattering is treated. Efficient expressions are derived for these scattering mechanisms, although reduction to a single integral was not possible like in the case of bound-to-bound transitions.

The evaluation of these expressions will be deferred to chapter 3.
Chapter 3

Transport modeling for Quantum Cascade Lasers

For the most part, QCL has historically progressed with very little theoretical input. The typical design process for lasers is exemplified by ref. [30], in which incremental changes are made to existing designs largely based upon the alleged “intuition” of the engineer. This approach has carried progress in THz QCL up to maximum operating temperatures of around 186 K. [24] My belief, however, is that this approach will no longer serve moving forward, and that computational design will play a larger role future design. Indeed, the approach to the present maximum of 200 K was achieved through careful numerical optimization by the researchers at NSERC. [1]

This chapter of the thesis details efforts to develop a useful model of QCL transport. An emphasis here is placed upon balancing efficiency versus accuracy. The justification is that a fast simulator that enables the engineer to quickly iterate through many variations of a design is superior to a marginally more accurate implementation that takes vastly more computational resources.

This thesis is primarily an experimental work (rather than computational or theoretical), and as such, little time could be afforded for programming and debugging. Therefore, to reduce development time, all programming was performed in MATLAB. MATLAB’s economy of syntax and wealth of numerical libraries made model prototyping far faster than working in a compiled language such as C/C++/Fortran.
3.1 Formalisms for transport

Broadly speaking, there exists three approaches to QCL transport, differentiated by the degree to which quantum mechanics is employed. At the highest level of accuracy are the fully quantum models of QCL transport, most notably embodied in the method of non-equilibrium Green’s functions (NEGF) (see, for example, [46, 54]). A general consensus exists that QCL transport happens through a combination of coherent and incoherent transport, and a fully quantum mechanical method such as NEGF is necessary to deal with these two processes on equal footing. While theoretically exact, the computational burden of NEGF and related methods tends to be severe, and so design with NEGF can be a very time consuming process.

In contrast, semiclassical models assume incoherent processes dominate. Electrons are modelled as scattering from subband to subband analogous to the dynamics of the Boltzmann transport equation in bulk materials. The only employment of quantum mechanics is in the determination of subband structure (eigenstates) and scattering rates (through Fermi’s golden rule). The most notable member of this family is Monte Carlo methods (see for example [48, 55]). However, such methods are unable to realistically treat tunneling effects. This can lead to unphysical results when localized eigenstates anticross and create coherences over unrealistically long distances (several QCL periods).

In between the fully quantum and semiclassical models are mixed models. In such models, typically a judgement needs to be made about whether coherent (tunneling) transport or incoherent (scattering) transport is dominant at different locations inside the laser, and one of these are treated accurately at the expense of the others. Examples of such models include the density matrix augmented Monte Carlo method of ref. [56], as well as the rate equations approach of ref. [57].
3.2 Choice of model and physical assumptions

At the beginning of this work, much time was devoted to the development of NEGF methods. Ultimately, however, this approach was abandoned due to its computational expense. In particular, studying electron-electron scattering was very difficult due to the presence of a large number of high-dimensional integrals. In pursuit of a faster model, a mixed model approach based on the work of Terazzi (ref. [37]) was implemented. The assumptions of the model are as follows.

- Subbands carrier distributions are assumed to obey Boltzmann statistics, and hence can be parameterized with a temperature $T$ and population density $n$ (or equivalently, with a total kinetic energy $K$ and chemical potential $\mu$). As shown in chapter 2, this assumption vastly simplifies the calculation of scattering rates. Such as an assumption implicitly assumes strong intrasubband scattering, such as through e-e interactions.

- The in-plane dispersion is characterized by a single effective mass $m_e$ for all subbands. In-plane non-parabolicity and anisoptropy are neglected. While questionable in midinfrared QCLs, this assumption works well in THz QCLs due to the low carrier densities and energies.

- A uniform 3D continuum exists starting at the barrier edges. All mismatches in quantum impedance are ignored.

- State-blocking is neglected. This is again justified by the typically low electron concentrations in THz QCLs. Further elaboration on the justification for neglecting state-blocking is found in section 7.

- A constant full-width-at-half-maximum (FWHM) of 1 THz is assumed for all intersubband optical transitions.

- Intermodule QCL transport is assumed to be dominated by resonant tunneling at the injector barrier, whereas intramodule transport is dominated by scatter-
ing. This is not a wholly justified assumption, but the injector anticrossing in
typical QCL designs tends to be the smallest.

- For the purposes of tunneling, a global pure dephasing time $T^*$ is used to esti-
mate dephasing by intrasubband scattering.

The last three points are likely the greatest weaknesses of the present model. The
assumption of constant bandwidth ignores all details about decoherence processes.
The division between inter- and intra-module transport renders the simulation unsuit-
able for unrestricted design optimization, because the internal barriers must remain
thin enough to assume strong coherence. The constant $T^*$ introduces an uncertain
fitting parameter, and moreover assumes that intrasubband based dephasing is tem-
perature independent; strictly speaking, there is no reason to believe that this is
true.

The core technique of the simulation relies on using the Kazarinov-Suri’s formalism
to reduce quantum tunneling to effective tunneling rates that stand on equal footing
with scattering rates calculated from Fermi’s golden rule. As such, this thesis will
refer to it as the effective rate equations (ERE) model.

### 3.3 Computational basis

This work employs the single-period eigenbasis of energy for transport calculations.
Ultimately, the Hamiltonian structure of figure 3-1 is sought-after, in which the on-
diagonal blocks describes the spectrum for intramodule scattering calculations, and
the off-diagonal blocks describe the tunneling matrix elements. Due to the need
for self-consistent solution of the Poisson equation and Schrodinger equation, these
eigenstates must be repeatedly calculated. Although the eigenstates can be calculated
directly from using the real-space techniques described in section 2.1, this method is
slow in that it requires diagonalization of a rather large matrix.

To improve the computational efficiency, the eigenbasis is instead calculated from
an intermediate basis. Several choices were tried. The most theoretically appealing
Figure 3-1: Schematic of the matrix structure of the Hamiltonian for transport simulations. The on-diagonal blocks describe the spectrum for intramodule scattering calculations, and the off-diagonal blocks describe the tunneling matrix elements.

choice was to use the Wannier states of the ideal superlattice and their associated Hamiltonian. The idea was to diagonalize the states located inside one module with the electric field and Poisson potentials added, then apply the transformation matrix to the off-diagonal blocks. This method was employed in ref. \[58\] for studying bound-to-continuum absorption. However, this method was found to yield states that were occasionally in poor agreement with real-space diagonalization. Particularly, the tunneling interactions were at times very different from the anticrossings calculated from direct diagonalization of the two-module real-space Hamiltonian.

Ultimately, an alternative approach was adopted, in which the intermediate basis is taken to be the flat-band tight binding basis of a single QCL period (see figure 3-2). Only states below the barriers at flatband are retained. Eigenstates were calculated from the intermediate basis by adding in the influence of the electric field and any poisson effects, and rediagonalizing. Tunneling interactions between states in adjacent QCL periods is then calculated according to the tight-binding method of Terazzi,
3.4 Computation of matrix elements for scattering

This section deals with the efficient calculation of matrix elements needed for scattering in Fermi’s golden rule. Although their calculation is straightforward, the need for self-consistent Schrodinger-Poisson iterations means that the matrix elements, like the wavefunctions themselves, need to be repeatedly evaluated. Direct evaluation of these matrix elements from their associate wavefunctions is slow due to the need for numerical integration. A more efficient method is to calculate the flatband matrix elements once at the beginning of the simulation, and apply basis transformations to obtain matrix elements in the presence of electric fields and band-bending.

As stated in chapter 2, the scattering mechanisms included in this thesis are longitudinal optical phonons (LO phonons), interface roughness, impurity, and electron-electron (e-e). Due to some mathematical similarities between LO phonon scattering...
and e-e scattering, calculations performed for the Coulombic e-e scattering can be reused. Interface roughness matrix elements are trivial due to the presence of the \( \delta \)-function in the perturbing Hamiltonian.

### 3.4.1 Coulombic form factors for electron-electron and LO phonon scattering

Central to the study of e-e scattering are the e-e form factors. \cite{39, 59} In the single band model, these are defined by

\[
F_{ijkl}(q) = \int dz \int dz' \phi_i^*(z) \phi_j(z) e^{-q|z-z'|} \phi_k^*(z') \phi_l(z')
\]

where the second line specializes to the case of real wavefunctions (typically true) and a simplified notation for the argument dependence of the wavefunctions is adopted. \( q \) is the magnitude of the exchanged momentum. Generalization to the multiband case is straightforward (e.g. \( \phi_i \phi_j \rightarrow \sum_\beta \phi_\beta \phi_\beta \), where the summation is over the bands).

The particular form factor \( F_{abab} \) is important for LO phonon scattering as well (see section 2.3.5).

Evaluating equation 3.1 through direct evaluation of a 2D integral is computationally inefficient. A superior method is given in ref. \cite{60}, in which Fourier decomposition of the wavefunctions (through the fast fourier transform) is employed to evaluate equation 3.1 without a 2D integral. More details can be found in refs. \cite{60} and \cite{61}.

However, even this spectral calculation can be time consuming if performed repeatedly. This section therefore focuses on deriving the \( F_{ijkl} \) from the flatband matrix elements \( F_{abcd} \).

Although \( F_{ijkl} \) is a 4-tensor, the index pairs \( ij \) and \( kl \) may be regarded as monolithic indexes. By doing so, \( F_{ijkl} = F_{ij,kl} \) is no more than a regular 2D matrix. Supposing that the eigenstates can be expanded in terms of the flatband wavefunctions
\( f \) as \( \phi_i = \sum_a c_{ai} f_a \), then the form-factor expands as

\[
F_{ijkl} = \int dz \int dz' \psi_i \psi_j e^{-q|z-z'|} \psi_k' \psi_l'
\]

\[
= \sum_{abcd} c_{ai} c_{bj} \left[ \int dz \int dz' f_a f_b e^{-q|z-z'|} f_c' f_d' \right] c_{ck} c_{dl}
\]

\[
= \sum_{ab} \sum_{cd} A_{ijab} F_{abcd} B_{cdkl}
\]

This takes the form of a matrix transformation.

Technically, the problem is solved at this point, but there is substantial redundancy in \( F_{ijkl} \) that can be eliminated for further gains in efficiency. Most notably, \( F_{ijkl} \) is unchanged by swapping the indices \( i \leftrightarrow j \) and \( k \leftrightarrow l \) \( (F_{ijkl} = F_{jikl} = F_{ijlk} = F_{jilk}) \). Therefore, a factor of \( \sim 4 \) savings in memory is possible by storing only unique pairs \( ij \) and \( lk \). However, this also requires a more complicated transformation.

Consider the matrix product \( \sum_{ab} A_{ijab} F_{abcd} \). It can be divided into terms where \( a = b \) and terms where \( a \neq b \).

\[
\sum_{ab} A_{ijab} F_{abcd}
\]

\[
\begin{align*}
&= \sum_a A_{ijaa} F_{aacd} + \sum_{a} \sum_{b=a+1}^{\infty} A_{ijab} F_{abcd} + \sum_{a} \sum_{b=a+1}^{\infty} A_{ijba} F_{bacd} \\
&= \sum_a A_{ijaa} F_{aacd} + \sum_{a} \sum_{b=a+1}^{\infty} (A_{ijab} + A_{ijba}) F_{abcd} \\
&= \sum_{a} \sum_{b=a}^{\infty} \tilde{A}_{ijab} F_{abcd} \\
&= \sum_{ab} \tilde{A}_{ijab} F_{abcd}
\end{align*}
\]

where the fact that \( F_{abcd} = F_{bacd} \) has been used. The tilde indicates a sum over unique pair indices, and the modified transformation matrix \( \tilde{A} \) is defined by

\[
\tilde{A}_{ijab} = \left(1 - \frac{1}{2} \delta_{ab}\right) (A_{ijab} + A_{ijba})
\]

\[
= \left(1 - \frac{1}{2} \delta_{ab}\right) (c_{ai} c_{bj} + c_{ib} c_{aj})
\]
By similar procedure, a modified matrix $\tilde{B}$ is defined by

\[
\tilde{B}_{cdkl} = \left(1 - \frac{1}{2}\delta_{cd}\right) \left(B_{cdkl} + B_{dckl}\right)
\]

\[
= \left(1 - \frac{1}{2}\delta_{cd}\right) \left(c_{ck}c_{dl} + c_{dk}c_{cl}\right)
\]

(3.5)

In terms of unique index pairs then, the transformation is given by

\[
F_{ijkl} = \tilde{\sum}_{ab} \tilde{\sum}_{cd} \tilde{A}_{ijab} F_{abcd} \tilde{B}_{cdkl}
\]

(3.6)

This section finishes with some comments on the analytic properties of $A$ and $B$ versus $\tilde{A}$ and $\tilde{B}$. Both pairs are tranposes; that is,

\[
A = B^T
\]

\[
\tilde{A} = \tilde{B}^T
\]

(3.7)

Furthermore, $A$ and $B$ are mutual inverses. This can be verified from their definitions (equation 3.2).

\[
(AB)_{ijkl} = \sum_{ab} A_{ijab} B_{abkl}
\]

\[
= \sum_{ab} c_{ai}c_{bj}c_{ak}c_{bl}
\]

\[
= \left(\sum_{a} c_{ai}c_{ak}\right) \left(\sum_{b} c_{bj}c_{bl}\right)
\]

\[
= \delta_{i,k}\delta_{j,l}
\]

\[
= \delta_{ij,kl}
\]

(3.8)

In contrast, $\tilde{A}$ and $\tilde{B}$ are not mutual inverses. It can be shown that

\[
\left(\tilde{A}\tilde{B}\right)_{ijkl} = \delta_{i,k}\delta_{j,l} + \delta_{i,l}\delta_{j,k} - \sum_{a} c_{ai}c_{aj}c_{ak}c_{al}
\]

(3.9)
3.4.2 Coulombic form factors for impurity scattering

Similar to electron-electron scattering, the form factor for impurity scattering is given by

\[ G_{ij}(z_\alpha, q) = \int dz \phi_i \phi_j e^{-q|z-z_\alpha|} \]  

(3.10)

where \( z_\alpha \) is the dopant location perpendicular to the plane of growth. Unlike the e-e form factors, no particular advantage was found in evaluation through Fourier decomposition. Equation 3.10 is evaluated directly using real-space integration. Transformation from the flatband form factors is straightforward, given by

\[ G_{ij}(z_\alpha, q) = \int dz \phi_i \phi_j e^{-q|z-z_\alpha|} = \sum_{ab} c_{ai} \left[ \int dz f_a f_b e^{-q|z-z_\alpha|} \right] c_{bj} \]  

(3.11)

3.4.3 Screening and the Coulombic matrix elements

Screening of Coulombic interactions is included through the use of an anisotropic Debye screening wavevector given by

\[ q_D = \sqrt{\frac{n_{3D} e^2}{\varepsilon k_B T_e}} \]  

(3.12)

where \( n_{3D} \) is the average volumetric electron concentration, \( T_e \) is an average electron temperature, and \( \varepsilon \) is the static permittivity. For a typical electron density of about \( n_{3D} = 6 \times 10^{15} \text{ cm}^{-3} \) at \( T_e = 200 \text{ K} \), \( q_D \sim 2 \times 10^7 \text{ m}^{-1} \). The corresponding Debye length is about 50 nm, comparable to a QCL period length.

Matrix elements for e-e scattering are hence given by

\[ W_{ijkl}(q) = \frac{e^2}{2A\varepsilon} \frac{F_{ijkl} \left( \sqrt{q^2 + q_D^2} \right)}{\sqrt{q^2 + q_D^2}} \]  

(3.13)
and matrix elements for impurity scattering are given by

$$Y_{ij}(q) = \frac{-e^2}{2\sqrt{q^2 + q_D^2}} \int dz_\alpha N_D(z_\alpha) G_{ij}(z_\alpha, \sqrt{q^2 + q_D^2})$$  \quad (3.14)$$

The evaluation of equation (3.14) requires an integration over the dopant distribution in addition to an integral implicitly lurking in the form factor $G_{ij}$. In the general case, this means that impurity matrix elements require a rather slow 2D integral to complete. However, this time can be significantly reduced by observing that in most cases, only a small portion of a THz QCL is doped.

More detailed descriptions of screening do exist, notably the Lindhart formulation of the Random Phase Approximation (RPA), but the computational expense was considered too high to merit their use. For example, Nelander has shown that isotropic Debye screening yields similar results to the static RPA, at least when the Debye length is comparable to the QCL period. Terazzi has also used the Debye screening model to reasonable effect in the description of Mid-IR QCLs.

A less well discussed issue is whether static screening is reasonable in THz QCLs. There is evidence to suggest that static screening systematically overestimates screening effects, and failure to include dynamic screening also excludes the possibility of collective electron scattering (electron-plasmon scattering). However, the inclusion of dynamic screening considerably complicates the mathematical description of Coulombic scattering. Given the low concentration of electrons in typical THz QCLs, plasmon effects should not be large (although perhaps even this point is disputed; see). Furthermore, screening is most important for intrasubband electron-electron scattering, to prevent a singularity in the mathematics of Fermis golden rule; for intersubband transitions, the rates are less affected.

Furthermore, the long Debye length of screening suggests that interperiod electron-electron interactions may be important, whereas only intraperiod scattering is presently included. However, due to the quartic scaling of the number of possible electron-electron interactions with the number of subbands, inclusion of even nearest neighbours quickly explodes into an intractable problem. Inclusion of nearest-neighbor
scattering would require some intelligent selection of a subset of possible interactions in future work.

3.5 Determination of populations

Subband and continuum populations are determined assuming that all temperatures are held fixed. Initially, no lasing is assumed, and scattering rates are calculated. The nonlinear rate equations (nonlinear because of electron-electron scattering) are linearized to get scattering time constants, whereupon the population rate equations form a linear algebraic system that can be solved using matrix inversion (supplemented by the conservation condition \( \sum_i n_i + N_{Cl} = n_{\text{total}} \), where \( l \) is the QCL module length). Mixing of the current previous solutions along with fixed point iteration is then use to converge to an answer of the nonlinear equations. This method is a bit crude and could use refinement, but it serves well enough.

If threshold gain is attained, then the the photon flux is increased and the populations are recalculated until the convergence criterion \( g_{\text{max}} - g_{\text{th}} = 0 \) is met. For this purpose, MATLAB’s fzero function is used.

3.6 Determination of subband temperatures

Subband temperatures are determined using the method of ref. [40]. In ref. [49], Slingerland also uses similar methods, although different expressions for kinetic energy transfer are used compared to chapter 2 of this thesis. With the subband populations held fixed, the subband temperatures are determined using a nonlinear root finding algorithm (fsolve in MATLAB). The criteria for convergence are the net rate of change of kinetic energy (KE) in each subband (determined by the heating and cooling rates calculated according to chapter 2). These rates are scaled by the lattice temperature and multiplied by an arbitrary user-chosen time constant to arrive at a dimensionless convergence criteria.

Initially, convergence problems were encountered with this method. The origins of
these problems were traced down to the fact that subbands with very low populations tended to have nearly zero KE rates regardless of the chosen set of temperatures. Therefore, a population threshold was implemented, below which the net KE rate of the associated subband would be ignored in determining convergence. This solution is not entirely satisfactory, in that it can lead to discontinuous jumps in subband temperature versus bias. On the other hand, such jumps only occur for nearly empty subbands anyway, which contribute little to the physics of the system.

In practice, determination of the subband temperatures tends to be one of the more time consuming steps. This computational expense is not generally worthwhile; as will be discussed in section 3.11, at high temperatures there is evidence to suggest that electron temperatures are essentially the same as the lattice.

3.6.1 Determination of the continuum temperature and mobility

The continuum temperature is determined differently, using a relaxation time approximation. Continuum electrons are assumed to be heated under the influence of the applied electric field $E$, and relax towards lattice equilibrium on some characteristic time scale $\tau_E$. The kinetic energy of an electron $E$ is assumed to be determined by

$$\frac{dE}{dt} = eE v - \frac{1}{\tau_E} \left( E - \frac{3}{2} k_B T_l \right) = 0 \quad (3.15)$$

where $v$ is the electron velocity in the continuum (itself a function of field). Assuming that the energy is related to the continuum temperature by $E = \frac{3}{2} k_B T_c$, then the latter is given by

$$k_B T_c = k_B T_l + \frac{2}{3} eE v \tau_E \quad (3.16)$$

The relaxation time $\tau_E$ is estimated as the ensemble average scattering rate of continuum electrons into bound subbands. It is also used to determine the low field mobility through

$$\mu = \frac{e \tau_E}{m_E} \quad (3.17)$$
and subsequently the continuum velocity through

\[ v = \frac{\mu \mathcal{E}}{\sqrt{1 + (\mu \mathcal{E}/v_{\text{sat}})^2}} \]  

(3.18)

The saturation velocity \( v_{\text{sat}} \) is chosen in the range of \( 1 \times 10^7 \, \text{cm/s} \) to \( 2 \times 10^7 \, \text{cm/s} \).

### 3.7 Band bending

Given an subband population \( n_i \) (units of inverse area), a uniform continuum distribution \( N_C \) (units of inverse volume), and doping distribution \( N_D(z) \), the charge distribution due to one QCL period is given by

\[ \rho_1(z) = -e \sum_i n_i \left( \sum_\beta |\phi_{\beta i}(z)|^2 \right) - eN_C + eN_D(z) \]  

(3.19)

This charge distribution is tiled to the left and right, in order to replicate the periodicity of the QCL. That is, the density

\[ \rho(z) = \rho_1(z - d) + \rho_1(z) + \rho_1(z + d) \]  

(3.20)

is calculated (\( d \) is the QCL period). This yields essentially three copies of the charge distribution, of which the central copy is chosen to represent the periodic charge distribution in the QCL. The periodicity of THz QCLs, at least at low temperatures (77K), is supported by recent scanning voltage-probe microscopy experiments by Dhar et al. [64].

Knowing this charge distribution, then Poisson’s equation

\[ \nabla \frac{1}{\varepsilon(z)} \nabla \varphi(z) = -\rho(z) \]  

(3.21)

is solve assuming Dirichlet boundary conditions for simplicity. Due to the periodicity of \( \rho(z) \), the choice of Dirichlet boundary conditions is equivalent to periodic boundary conditions (\( \varphi(z) = \varphi(z + L) \)) up to some physically unimportant constant term.
### 3.8 Hot phonon effects

There is some suggestion in the literature that hot (i.e., non-thermally distributed) phonons play a relevant role in QCL transport.\[65\] Both depopulation of the electrons from the lower laser levels and nonradiative leakage of upper level electrons to the lower laser levels at elevated temperatures results in copious LO phonon production, so this is not surprising. The existence of hot-phonons in THz QCLs is also supported by experimental results by Vitiello et al.\[66\]

To estimate these effects without accounting for the complete 3D phonon spectrum, the method of Lu and Cao is used.\[65\] For every pair of subbands, a 2D occupation factor $N_{LO} = N_{LO}(q)$ is tracked. This occupation factor is used solely for determining the LO phonon scattering between those two subbands. A Boltzmann-like transport equation can be written for the phonon occupation factor in the relaxation time approximation. Assuming phonon emission and absorption due to transitions between two subbands $a$ and $b$, this reads

$$\frac{dN_{LO}(q)}{dt} = 2 \sum_{\kappa_a} W_{em}(a\kappa_a \rightarrow b, \kappa_a - q) f_a(\kappa_a) - 2 \sum_{\kappa_a} W_{abs}(a\kappa_a \rightarrow b, \kappa_a + q) f_a(\kappa_a)$$

$$- \frac{N_{LO}(q) - N_{LO,th}}{\tau} \quad (3.22)$$

where $N_{LO}(q)$ is assumed to relax to the thermal expression $N_{LO,th}$ with time $\tau$ by decay into acoustic phonons (typically chosen in the range of 4-7 ps). The emission and absorption rates can be evaluated using the same mathematics as section 2.3.2. For example, for the emission rate, this yields

$$2 \sum_{\kappa_a} W_{em}(a\kappa_a \rightarrow b, \kappa_a - q) f_a(\kappa_a)$$

$$= \frac{1}{\hbar^2 A} |U(q)|^2 \left( \frac{n_a}{q} \right) \sqrt{\frac{2\pi m_e}{\kappa_B T_a}} \exp\left( -\frac{\epsilon_{\alpha}^e(q)}{\kappa_B T_a} \right) \quad (3.23)$$
where

\[ |U(q)|^2 = \frac{e^2\hbar\omega_{LO}}{4A} \left( \frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_s} \right) (N_{LO}(q) + 1) \left[ \frac{F_{ab}(q)}{q} \right] \]

\[ \epsilon'_a(q) = \left( \frac{E_b - E_a + \hbar\omega_{LO} + \epsilon_q}{4\epsilon_q} \right)^2, \epsilon_q = \frac{\hbar^2k^2}{2m_e} \]  (3.24)

Since a solution for \( N_{LO}(q) \) is ultimately sought-after, the emission term is better rewritten as \((N_{LO}(q) + 1)W_{em}(q)\). Similarly, the absorption term can be written in the form \( N_{LO}(q)W_{abs}(q) \). Therefore, equation 3.22 becomes

\[ \frac{dN_{LO}(q)}{dt} = (N_{LO}(q) + 1)W_{em}(q) - N_{LO}(q)W_{ab}(q) - N_{LO}(q) - \frac{N_{LO}(q) - N_{LO,th}}{\tau} \]  (3.25)

The steady state solution for the phonon occupation is therefore

\[ N_{LO}(q) = \frac{\tau W_{em}(q) + N_{LO,th}}{(1 - \tau W_{em}(q) + \tau W_{ab}(q))} \]  (3.26)

For example, at design bias and 150 K lattice temperature, typical maximum values of \( N_{LO}(q) \) are in the range of 0.1 to 0.3, depending on the transition. In ref. [66], \( N_{LO} \) of about 0.4 is measured under similar conditions. While the hot phonon effects appear to be somewhat underestimated in this work, the values are reasonable.

### 3.9 Overall Procedure

A simplified flowchart detailing major steps in the ERE is summarized in figure 3-3. Once the subband populations and temperature (and photon flux, if lasing) have converged, other information about the laser, such as current density, can be extracted.

### 3.10 Model validation

In the section, the validity of the ERE model is checked by testing it against data from historical QCLs. In addition, conservation of energy is demonstrated as proof
Calculate band-structure, flatband matrix elements

Initial guess for distribution (usually thermal)

Update continuum temp.

Update subband/continuum pop.

Threshold reached?

Y

Estimate photon flux (lasing), update pop.

N

Update Poisson potential

Update subband temp.

Update continuum $\mu, \tau$

Update phonon distribution

Pop, Temp, Flux converged?

N

Y

Return current, gain, etc.

Figure 3-3: Flowchart detailing major steps of QCL simulation.
of the validity of the thermal modeling used in the effective rate equations approach.

### 3.10.1 OWI222G

OWI222G (wafer number VB0240) is a previous record holder for highest $T_{\text{max}}$.\cite{24} Its band diagram is shown in figure 3-4. It was previously studied in ref.\cite{67} using the Monte Carlo density matrix (MCDM) methods of ref.\cite{56}, in the context of demonstrating the importance of lasing to the explanation of experimental transport data. In figure 3-5, the ERE calculated low temperature $IV$ is compared to the MCDM calculated $IV$ from ref.\cite{67}, as well as experimental data. For the ERE calculated $IV$, the following parameters are assumed: $T^* = 0.15$ ps pure dephasing; gain threshold of $23$ cm$^{-1}$; subband heating of $100$ K above lattice.

As seen in figure 3-5, MCDM and ERE give comparable results, with MCDM slightly over estimating all currents in general, and ERE overestimating the sub-threshold knee. Although both simulations methods agree reasonably well with experiment, the MCDM simulation took over a day to run, without doing self consistent Schrodinger-Poisson iterations. In contrast, ERE assisted by parallelization in MATLAB uses only about half an hour, even when doing Schrodinger-Poisson iterations.
Figure 3-5: OWI222G IV curves at 10 K calculated using a mixed Monte Carlo Density Matrix (MCDM) technique, and using the effective rate equations (ERE) approach. Both approaches agree reasonable well with experiment after a voltage translation to account for Schottky contacts (+7 V for MCDM, +5 V for ERE).
3.10.2 FL183S

FL183S is widely used in the THz-MMW group at MIT for THz QCL applications, particularly in the production of distributed feedback lasers. A four-well gain medium grown in GaAs/Al$_{0.15}$Ga$_{0.85}$As, it features an attractive combination of high electrical stability, reasonable current density, good $T_{\text{max}}$, and good power. It has been grown in multiple wafers with reproducible results (a testimony also to the high quality of growth from Sandia National Laboratories), and hence serves as a good reference for model calibration.

Surprising problems were encountered in the analysis of FL183S. Initial analysis indicated that FL183S has extremely poor gain, in contrast to its good performance in applications. The origins of this discrepancy were traced down to the presence of a high energy parasitic subband. Figure 3-6a shows the one module bandstructure of FL183S. The energy separation $E_{65}$ is very close to the lasing energy $E_{54}$. Because subband 6 is nearly empty due to its high energy, and subband 5 is highly populated by resonant injection, there is strong $5 \rightarrow 6$ absorption feature that drastically reduces the gain due to the $5 \rightarrow (4,3)$ lasing transition. Figure 3-6b shows the unclamped gain spectrum calculated with all subbands and calculated using only the first 5 subbands.

With the $5 \rightarrow 6$ absorption removed, good agreement with experimental results is achieved. This is illustrated in figure 3-7 which compares the computed and experimentally measured IV curves. For ERE simulations, the following parameters are assumed: $T^* = 0.15$ ps pure dephasing; gain threshold of 18 cm$^{-1}$; subband heating of 100 K above lattice. Apparent absence of $5 \rightarrow 6$ indicates that the present model has some shortcomings in its physics. For example, one possibility is that the high energy subband 6 has a lifetime so short that the $5 \rightarrow 6$ transition is broadened to the point of irrelevance, or that is its high energy makes it poorly determined in the tight-binding approach.
Figure 3-6: (a) Band diagram of FL183S. A spurious absorption feature appears in one-module simulations due to transitions between upper laser level 5 and high energy parasitic 6. (b) Influence of spurious absorption on gain spectrum: gain is reduced compared to the case when subband 6 is omitted from spectrum.
Figure 3-7: FL183S $IV$ curve computed at 10 K using ERE approach, in comparison to experimental data.
3.10.3 Conservation of energy

A key feature of the ERE approach is its ability to predict subband temperatures. To test the validity of this thermal modeling, conservation of energy is considered here.

From a thermodynamic viewpoint, a QCL under operation at electric field $\mathcal{E}$ and current density $J$ must dissipate energy at a rate $J\mathcal{E}$ (units of W/m$^{-3}$). Of the scattering mechanisms included in this thesis, only LO phonon scattering exchanges energy with the lattice. The energy dissipation rate is given by

$$\frac{\hbar \omega_{LO}}{\ell} \sum_{ab} (W_{LO,em}(a \rightarrow b) - W_{LO,ab}(a \rightarrow b)) \quad (3.27)$$

Equation 3.27 should equal $J\mathcal{E}$. Testing on several structures at design bias indicates an agreement of better than $\sim 0.5\%$.

3.11 Regarding electron heating

THz QCLs operate at very high electric fields. In a bulk material, such fields would lead to heat electrons to temperatures far higher than the lattice. This begs the question: are electrons in THz QCLs significantly heated above the lattice? The relevance of this question is that electron heating would help explain part of the mystery of why THz QCLs fail to operate at elevated temperatures. In this regard, the most crucial temperature is the temperature of the upper laser level, which determines LO phonon mediated scattering to the lower laser levels with exponential sensitivity.

The best experimental measurements of subband temperatures at present are based on microprobe photoluminescence (MPPL) experiments, most notably performed by Vitiello et al. [68] In a 5-level resonant phonon design at $T_L = 100$ K, electron heating $\Delta T$ was measured to be $\sim 100$ K for the upper lasing level, and insignificant for lower subbands. More recent studies on scattering assisted gain media measured $\Delta T \sim 40 - 50$ K or so. [69]

This section uses the intersubband thermal modeling of section 3.6 to study this
problem. Similar work has been performed by Han \cite{70} and Slingerland \cite{49}. Figure 3-8 shows the calculated electron subband temperatures for the four principle subbands of the record-holding $T_{\text{max}} \sim 200$ K design of ref. \cite{1}. At low temperature, there is significant electron heating due to the dominance of elastic scatters. However, as the lattice temperature rises, LO phonon scattering becomes stronger, and relaxes the subband temperatures. In the case of the upper laser level, below $T_L \sim 100$ K, the subband temperature is clamped to about 100 K. Above $T_L \sim 100$ K, the electronic temperature is only minimally heated above the lattice. This work agrees with ref. \cite{70} and \cite{49}. It is also consistent with transport simulation of mid-infrared QCLs by Terazzi, which were found better agreement with experiment when electronic and lattice temperatures were assumed to be equal. \cite{37}. The flatness of the electron temperature up until about $\sim 100$ K perhaps also explains the experimentally observed low temperature insensitivity of the threshold current $J_{\text{th}}$. The conclusion is that electron heating is essentially a low temperature phenomenon. This is qualitatively consistent with MPPL experiments, which are performed under cryogenic cooling to heatsink temperatures of about 50 K (the lattice is apparently warmer, at $\sim 100$ K). The implications for THz QCL design is that electron heating is not relevant for explaining the failure of lasing at high temperatures. However, it should be considered for modeling low temperature $IV$ curves. As the subband temperature modeling process is computationally expensive, my recommendation to the laser engineer is to calculate the heating at the design bias, and use the obtained value of heating for the computation of the rest of the $IV$. This is the appproach used in sections 3.10.1 and 3.10.2 to good effect.

Quantitatively, however, the resolution of subband heating is as yet incomplete. On theoretical grounds, electron heating is caused primarily by the elastic scattering of electrons from a high energy subband into a lower energy subband, with the intersubband gap energy manifesting as kinetic energy in the lower subband. Resonant tunneling should theoretically cause little heating. As such, the subband thermal modeling of section 3.6 predicts the lower laser levels to be the hottest, and upper laser level to be the coldest. In contrast, refs. \cite{68} and \cite{69} reach the opposite conclu-
Figure 3-8: Temperatures of the 4 lowest energy subbands in the record-holding $\sim$200 K design of ref. [1] at design bias. Significant electron heating is observed only at low temperatures. Electron temperature is mostly flat up until about $\sim$ 100 K, perhaps explaining the insensitivity of threshold current $J_{th}$ at low temperatures.
sion: both studies measure the greatest heating in subbands populated by resonant tunneling, with lower subbands being essentially in equilibrium with the lattice. Furthermore, ref. [68] measures significant heating even at a lattice temperature of 100 K, in contrast to the results here.

While MPPL measurements have provided valuable insight for discussion, temperatures are inferred from MPPL experiments using a rather involved fitting processes using a large number of parameters. Assessing the precision of the measured temperatures is somewhat difficult. More work is required.

3.11.1 Justification for the neglect of state-blocking

A key assumption of this thesis is that state-blocking (Pauli exclusion) is negligible. Quantitative justification is given here.

The chemical potential of a 2D subband is given by

$$\frac{\mu}{k_B T} = \ln \left( \exp \left( \frac{n_{2D} \pi \hbar^2}{m_e k_B T e} \right) - 1 \right)$$

(3.28)

and the maximum in the Fermi-Dirac occupancy is given by

$$f = \frac{1}{e^{-\mu/k_B T} + 1}$$

(3.29)

As seen in section 3.11, the electron temperature assumes a minimum temperature of about 100 K at design bias. For a QCL of typical doping, assuming that the entire population of $\sim 3 \times 10^{10}$ cm$^{-2}$ lies in a single subband, the maximum occupancy at 100 K is 0.12. Therefore state-blocking terms ($1 - f$ terms) are not that important (this may change as doping increases).

3.12 Model limitations

Despite its success at modeling low temperature transport, the ERE model developed in this chapter is still incomplete. For example, one such shortcoming is detailed in
In section 3.10.2, the gain spectrum in FL183S is shown to be contaminated by a spurious absorption feature caused by limitations in the tight-binding approach to basis state calculation. More seriously however, the ERE model is as yet unable to explain the temperature performance of any device. For example the $T_{\text{max}} \sim 200$ K device of ref. [1] is predicted to still possess $\sim 32 \text{ cm}^{-1}$ of gain at 200 K, compared to $53 \text{ cm}^{-1}$ at 4 K.

The reason for this failure is as yet unknown. Because current is underestimated at high temperatures, a likely possibility is that the upper-level lifetime is overestimated. This is as opposed to, say, a loss of coherence in resonant tunneling, which would decrease current. Nelander has also suggested gain broadening with increasing temperature as a possibility for loss of peak gain. [63] But while linewidth broadening would explain the loss of gain, it does not explain the increases in current versus temperature. Furthermore, there is some experimental evidence to the contrary, that intersubband THz line widths are fairly stable with regard to temperature. [71]

On the topic of coherence, there is a peculiarity in the choice of $T^* = 0.15 \text{ ps}$ and 1 THz FWHM for optical transitions. A 1 THz FWHM corresponds to a dephasing rate of about 0.3 ps. Since optical linewidth and tunneling have a common origin in quantum coherence, this begs the question why would tunneling dephase so much faster than optical coherence? There is some justification for this in that dephasing is a sensitive function of correlation (more correlation implying less dephasing). Since most optical transitions are essentially intrawell, and tunneling processes are by definition interwell, the scattering in the former case can be reasonably expected to be more correlated.

### 3.13 Influence of continuum interactions

If an overly long upper level lifetime is the reason for disagreement with experimental at high temperatures, then one must ask why? If the scattering mechanisms modeled are correctly treated, then there must be something missing. One possibility is the leakage of carriers to continuum states. The desire to study leakage to continuum is
what motivated the developments of sections 2.6 and 2.7

The relevance of leakage to continuum is not immediately obvious. The upper laser level is typically separated from the tops of the barriers by at least 50-70 meV, which is substantially greater than $k_B T$ even at 300 K. Given such a high thermal barrier, a large number of continuum electrons is not expected to exist in continuum states, at least not based on single body scatterers.

However, because THz QCLs typically operate at very high electric fields (10 kV/cm or more), some curiosity arose as to whether impact ionization could mediate the leakage of electrons to continuum. Whereas the energy threshold for impact ionization in bulk semiconductors is essentially determined by the material bandgap, the threshold in an intersubband process is expected to be the distance from the edges of the bound subbands to the top of the barriers. This latter energy is much smaller than the former, thus one might postulate that the continuum electrons in QCL, accelerated under the influence of the electrical bias, might be energetic enough to liberate a significant number of electrons from the bound subbands. A small number of initial electrons can hence liberate a larger number due to the avalanche effect.

To clarify, “avalanche” here is not meant in the same sense as the true avalanche process in an avalanche photodiode. In the latter, the number of carriers continues to grow throughout the length of the device. In a avalanche photodiode, this exponential-in-distance growth of electrons would be balanced by the same growth of holes in the opposite direction, thus ensuring current conservation. A THz QCL, however, is unipolar device; in the absence of hole currents, a true avalanche process would result in band-bending from one QCL period to the next, violating periodicity. But as previously mentioned, the periodicity of QCLs has been verified experimentally, at least at low temperatures ($\sim 77$ K). [64]

As it turns out, inclusion of impact ionization had little effect. Although intersubband impact ionization could be faster than the reverse process of intersubband Auger recombination for states very close to the top of the barriers, it was still a slower process than electron capture through LO phonon emission. Scattering rates of about $\sim 1$ ps$^{-1}$ per subband were found, meaning overall average capture times of
$\sim 0.2$ ps. Even at 300 K, only around 2% of the total electron population is predicted to exist in continuum states.

In summary, all simulations indicate a very low percentage of electrons in the continuum. This is consistent with experimental results from [7.2] in which only minor improvements at best are shown when using taller quantum barriers.

### 3.13.1 An alternative treatment of LO phonon capture

The absence of significant impact ionization effects is due to the extremely fast LO phonon captures times. One possible objection to the approach of this thesis is the assumption of a thermalized, shifted Maxwellian distribution in the continuum. The existence of thermalization amongst above-barriers states is uncertain. Therefore, this section addresses whether a more generalized distribution would make difference.

In more detail, the question is whether continuum electrons could possibly be ballistic on lengths scales longer than a QCL period length, and hence gain energy moving from one period to the next. The form of such a distribution is uncertain, and would likely not be periodic anyway. However, for this to occur, such high energy electrons must exhibit significantly lower probability of capture compared to electrons near the barrier tops. Therefore, independent of assumptions regarding the continuum distribution, the carrier lifetime in the continuum is useful quantity to examine.

"Lifetime" per se is not well defined for transitions from 3D to 2D; capture is instead described by a *scattering velocity* $v_{e-LO}$ defined such that the continuum to bound subband transfer rate is given by

$$W (C \to b) = \frac{1}{V} \sum_{\overrightarrow{K}_a} f_C (\overrightarrow{K}_a) v_{e-LO} (\overrightarrow{K}_a)$$

The momenta of electrons accelerated in an electric field from zero energy should be $\hat{z}$ dominant. Thus, consider the case of vanishing lateral momentum in the continuum ($k_a = 0$, $\overrightarrow{K}_a = k_{za} \hat{z}$). Under this condition, the methods of section 2.6.1 can be used
to show that \( v_{e-LO} \) as a function of kinetic energy \( \epsilon_a = \frac{\hbar^2 k_{za}^2}{2m_e} \) is given by

\[
v_{e-LO}(\epsilon_a) = \frac{m_e}{2\pi \hbar^3} \frac{e^2 \hbar \omega_{LO}}{2} \left( \frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) (N_{LO} + 1) \int dk_{zb} |\Phi_b(k_{zb})|^2 \frac{q^2 + (k_{zb} - k_{za})^2}{\left( q^2 + (k_{zb} - k_{za})^2 + q_s^2 \right)^2} \Theta \left( \frac{\hbar^2 k_{za}^2}{2m_e} > \hbar \omega_{LO} - \Delta E_c + E_b \right)
\] (3.31)

Figure 3-9 shows the scattering velocity calculated for the 6 bound subbands of the design of [1] (this design was also used in section 3.11). As seen in figure 3-9, the scattering velocities are not a strong function of energy, and are moreover comparable to reasonable estimates of the electron drift velocity in the continuum. Therefore, even accounting for more arbitrary distributions, LO phonon scattering remains a ultra-fast process, at least based on this formalism. The possibility of quasi-ballistic electrons seems low.

### 3.14 Conclusions

This chapter is closely aligned with chapter 2 motivating their side-by-side placement in the ordering of this thesis. However, chronologically the work of this chapter was done after the work of chapter 6. The model was built in response to a perceived need for better modeling tools to enable future design, particularly for the handling of Poisson effects.

To that effect, an effective rate equations model for QCL transport was developed. It is built almost entirely from first principles, with most input parameters reasonably well defined. The major uncertain parameters needed for fitting are the pure dephasing time \( T^* \) for tunneling, the optical linewidth, and — to a lesser extent—the threshold gain for lasing. Low temperature \( IV \) curves are predicted reasonably well, but the model fails at predicting temperature performance.

But while imperfect, my hope is that this tool will be nonetheless be a useful
Figure 3-9: Scattering velocities for electron capture from continuum via LO phonon emission. The THz QCL design is that of ref. [1]. Curves are evaluated according to expression 3.31.
addition to the toolkit of the laser engineer. It can diagnose whether lasing is possible in new designs, whether the threat of early negative differential resistance is a possibility, and it can estimate band-bending effects.

Given also the similarities of this work and the work of Terazzi, [37] a summary of the main differences is also given here.

- Ref. [37] includes explicit calculation of intraband dephasing times and optical linewidths. Said calculations worked well for MIR QCLs, but failed for THz QCLs. Therefore, this work treats both quantities as fitting parameters.

- Ref. [37] uses a global electron temperature, either in equilibrium with the lattice or through the determination of a global kinetic energy balance. This work explicitly calculates individual subband temperatures through a detailed kinetic energy subband for each subband.

- Ref. [37] excludes e-e scattering. This work includes detailed e-e scattering.

- Ref. [37] models only transitions between bound subbands. This work makes a (crude) attempt to include continuum states, although these are found to have little effect.
Chapter 4

Experimental methods and measures

This chapter summarizes the experimental methods used in the fabrication and measurement of THz quantum cascade lasers.

4.1 Making Terahertz Quantum Cascade Lasers

4.1.1 Gain medium growth

THz QCL gain media of this thesis were grown using molecular beam epitaxy (MBE) by Dr. John Reno and his staff at Sandia National Laboratories (SNL). Growth is done at a substrate temperature of 630°C. Starting with a bare GaAs substrate (either n+ or semi-insulating), growth begins with a 300 nm epitaxial GaAs buffer; a 400 nm AlxGa1−xAs etch stop layer (x ~ 55%) for substrate removal later; and a 100 nm heavily doped n+ layer for Schottky contact formation. Following these three layers, the gain medium proper is grown. QCL periods are stacked to a height of 10 µm; since each QCL period is typically around 50 nm long, there is typically around 200 repeats. The QCL growth is then followed with a “half period” consisting of just the QCL injector region; a 50 nm heavily doped n+ layer for Schottky contact formation; and a thin 10 nm layer of extremely highly doped (5×10^{19} cm^{-3}) GaAs.
The substrate temperature is then reduced to 250°C, and topped with an extremely thin 3.5 nm layer of low temperature grown (LTG) GaAs. The LTG GaAs is riddled with defect states, which forms a tunneling-based ohmic contact at the interface. Elaboration on this technique is given in refs. [72, 73].

The final product is depicted in figure 4-1. A sample growth sheet is given in appendix B.

4.1.2 Procedure for double-metal waveguide THz QCLs

This section describes the clean room fabrication procedures for fabricating THz QCL laser dies starting from the bare MBE growth of section 4.1.1. The overall fabrication process is shown in figure 4-2, and the main steps are sketched here. Some troubleshooting advice is documented in appendix C.

From the received MBE wafer, samples of typical size 1.2×1.7 cm$^2$ are cleaved. In addition to the cleaved MBE sample, a slightly larger n$^+$ substrate is necessary; this substrate is dubbed the “receptor,” as the MBE growth will be transferred to it during the fabrication process.

Both the MBE sample and receptor substrate are metal coated (typically at 1 to 2 × 10$^{-6}$ torr vacuum, at 1 Å/s). The MBE sample and receptor are bonded together using metal-metal thermocompression at 300°C for 1 hour in vacuum, followed by a 1 hour anneal at the same temperature in nitrogen. [74] A typical metal stack is 10 nm Ta/250 nm Au, but many other metal stacks have been employed in this thesis, including Ti/Au, W/Ag, Ta/Cu, and Ta/Au/Ta/Cu. In the Ta/Au/Ta/Cu stack, the
Figure 4-2: Major steps in double-metal waveguide formation for Terahertz Quantum Cascade laser fabrication. Steps read top-to-bottom, left-to-right.
copper has no waveguiding function, but merely acts to mechanically bond the MBE sample and receptor substrate together. The same bonding recipe gives good results for Au-Au, Ag-Ag, and Cu-Cu thermocompression bonding; a controlled study has not been performed, but anecdotally, Cu-Cu bonding appears to yield fewer bonding defects compared to Au-Au and Ag-Ag. Graphite spacers are useful for distributing the force during bonding.

Following wafer bonding, the native MBE substrate is lapped down to a thickness of \( \sim 70 \) to 100 \( \mu \text{m} \). This remaining MBE substrate is then removed by wet etching in a 3:1 (by volume) solution of citric acid (1 g/mL), 20% hydrogen peroxide (H\(_2\)O\(_2\)), which stops selectively at the AlGaAs etch stop. The receptor substrate backside is protected by photoresist during this process. Two important notes are in order here. First, the etching solution ages with time, and needs to be mixed fresh at roughly 1 hour intervals (I thank Mr. Sylvain Laframboise, formerly of NRC Canada, for this information). Second, Ag is an extremely efficient catalyst for H\(_2\)O\(_2\) decomposition, and thus rapidly inactivates most GaAs wet etches. Any exposed Ag needs to be carefully masked with photoresist, which makes the fabrication of Ag waveguides especially challenging. The exposed etch-stop layer is then removed in concentrated (49%) hydrofluoric acid (HF). This completes the transfer of the gain medium epitaxial layer to the receptor.

Top contact/waveguide metal is then defined by image reversal photolithography. The final top metallization layer is always Au. Mesa definition then occurs through dry-etching in 0.5/3/16 sccm Cl\(_2\) : SiCl\(_4\) : Ar, for which top Au acts as a self-aligned hard mask for etching. I thank Dr. Tsung-Yu Kao for the development of the dry etch recipe used in this thesis. \([75]\) The dry etch recipe achieves vertical sidewalls through the growth of a passivating Si-based layer during the etch. This layer must be removed to facilitate successful device cleaving. Sidewall passivation removal can be achieved etching in an isotropic SF\(_6\) plasma, \([75]\) or by immersion in buffered-oxide etchant (BOE). If the sample is not BOE safe (for example, it contains Ti metallization), then a glycerol containing BOE formulation such as Transene pad-etchant can be used instead of regular BOE.
Alternatively, a second photolithography step may be used to define a photoresist soft mask for wet etching. The typical etching solution used is a 1:1:25 (by volume) solution of 20% H$_2$O$_2$, concentrated (85%) phosphoric acid, and deionized water. This etchant yields an extremely smooth GaAs etch with a reproducible etch rate of approximately 0.25 $\mu$m per minute. However, the phosphoric acid etchant suffers badly from galvanic enhanced etching, and must therefore be kept well separated from metal. For example, etching should not proceed all the way down to the bottom metal, lest the mesa get undercut by lateral etching.

Finally, the receptor substrate is lapped down to $\sim 200 \mu$m by lapping, and coated with backside Ti/Au layer for soldering. This thinning step is not strictly necessary, but eases wafer cleaving and heat removal from the laser die during experiments.

A photograph of a finished laser die is included in figure 4-3a. The long lines correspond to Fabry-Perot laser bars of varying widths. A SEM of a 40 $\mu$m wide bar is shown in figure 4-3b.

### 4.2 Device mounting

On the finished laser die, Fabry-Perot laser cavity widths are defined lithographically. The cavity lengths are defined by wafer cleaving. The die is held in place by a soft material such as a cleanroom fabrication wipe and nicked by a diamond scribe. It is then turned face-down onto another wipe, and pressure is applied to the location of the scribe mark to initiate the cleave. This procedure is repeated as necessary to define a smaller die with the appropriate laser cavity lengths.

The cleaved die is then die-bonded to a copper substrate using indium soldering. To start, the copper substrate surface is lightly sanded with a fine paper (9 $\mu$m to 30 $\mu$m particle) size) to smoothen the surface, and cleaned with isopropyl alcohol (IPA). A small piece of indium foil ($\sim 0.010$” thickness) is cut to a slightly larger size than the cleaved die.

The second step is to bond the foil to the copper substrate. Any residual metallic oxides must be removed from both the copper and indium prior to bonding. This
Figure 4-3: (a) Optical microscope image of the mesa pattern on a finished laser die. (b) Scanning electron microscope image of cleaved laser facet (mesa height is 10 µm, for scale).
is accomplished by wiping metallic surfaces with an acidic soldering flux (HCl with addition of ZnCl). In quick succession, the copper surface is wiped, then one side of the indium foil, and then the two cleaned surfaces are pressed together (a glass slide is useful for handling the indium foil). This step must be done quickly (within seconds, preferably), else not only will the surface oxide reform, the acid flux itself will react to form a whitish surface dross. If done sufficiently quickly, the two metal surfaces will adhere immediately.

The third step is placement of the laser die. The opposite surface of the indium is cleaned with IPA. Then in quick succession, it is wiped with acid flux and the laser die is positioned on top. Again, the acid-flux should not be left standing for significant periods of time, and the clean indium surface will adhere to the bottom gold of the laser die almost immediately. This step is the trickiest part of mounting, as it requires both precision and speed.

The fourth step is the actual soldering step. It should be done immediately after the third step (although response times measured in seconds are not necessary). The substrate and die are placed together on top of a chip bonder hot plate set just hot enough to melt indium. The bonder arm applies pressure to keep the die in place while the indium melts (if die positioning was not entirely satisfactory during the third step, some repositioning of the die is possible at this point). Once the indium melts, the hot plate heat is switched off, and the soldering bond is left to solidify. During the solidification period, the die must continue to be held in place; if pressure is released, the die will start to float on top of the liquid indium.

Finally, electrical connections are established to the laser bars using gold-wire bonds. GaAs is a much more fragile semiconductor than Si, so care must be taken not to accidentally crush or fracture the laser bar while bonding. Minimum ultrasonic powers and times should be applied.

The device is now ready for measurement. A photograph of a bonded laser die is shown in figure 4-1.
4.3 Characterization of Terahertz Quantum Cascade Lasers

This section details the characterization methods of THz QCLs.

4.3.1 Cryogenic measurements

The QCL devices were mounted in closed cycle pulse-tubed helium cooler for measurement (model PT810, built by Cryomech Inc.), which cools down to $\sim 8$ K. Typically, two devices of the same fabrication were mounted and measured to reduce the likelihood of statistical variation within the die. For example, a particular device might just happen to contain a bonding or growth defect, and consequently have inferior performance.

During the cool down from room temperature down to 8 K, the maximum lasing temperature would be measured for both devices. The device with the higher $T_{\text{max}}$ would be selected for further characterization. That said, the cumulative improvements in processing and measurement techniques made over the course of this thesis have made intra-die statistical variation very small. Typically, two devices will have a $T_{\text{max}}$ that agrees to within $\sim 2$ K of one another, often less. Repeated measurements
of the same die in hopes of getting a superior $T_{\text{max}}$ at the margins of the probability distribution were once standard, but are at present rarely worth the effort.

### 4.3.2 Electrical and Optical Detection

Empirically, larger devices tend to have superior $T_{\text{max}}$. Therefore, gain medium characterization was typically performed using devices of approximate dimensions 2 mm $\times$ 150 $\mu$m. Such lasers were typically biased to 15-20 V and drew 6-8 A of current. THz QCLs at present have less than 1% wall-plug efficiency, so all electrical power is essentially dissipated as heat. Due to this tremendous heat generation, such devices cannot be biased continuously without device damage. (Continuous-wave (CW) operation typically requires devices in the 20-40 $\mu$m width range). The large devices of this thesis were therefore characterized exclusively in pulsed mode.

Lasers are biased using a Avtech model AVR-250B pulsed supply. Typical pulsing conditions were 400 ns at 300-1000 kHz. These extremely low duty cycles (0.012% to 0.04%) are necessary to avoid device heating. Early literature worried even about device heating over the course of of a single pulse, and thus chose much shorter pulse lengths, typically around 100 ns. However, such short pulses tend not to be square without special impedance matching measures, and thus distort electrical measurements. The desire for clean electrical measurements motivates the choice of a longer pulse in this thesis; empirically, no effects on $T_{\text{max}}$ has been observed (this contrasts early literature, perhaps indicating an improvement in fabrication and mounting methods),

Voltage and current were monitored using two box-car averagers. Voltage measurements were made electrically parallel to the device under test; current was indirectly sampled using a inductively coupled current sensor.

For optical detection, two instruments were used. For absolute power measurements, a calibrated power meter developed by Thomas Keating Ltd. was used. The average power measured by the meter is then divided by the duty cycle to obtain the peak pulsed power.

For $LI$ and spectral measurements, the optical power output was measured using
a liquid helium (LHe) cooled Ga-doped Ge photodetector, built by Infrared Laboratories Inc. Detector noise can mask a small optical signal, so the use of low-noise detectors is essential for measuring the highest possible $T_{\text{max}}$; this justifies the use of a LHe cooled detector instead of cheaper pyroelectric detectors. Prior to recording, the output of the photodetector is fed through a lock-in amplifier locked to the bias repetition rate. The uncalibrated power measurement is then scaled to match the absolute power measured by the Thomas Keating meter.

The Ge:Ga photodetector is an excellent instrument with very low noise and moderately fast response times. Aside from the monetary expense of requiring liquid helium, the main disadvantage of the Ge:Ga photodetector is that it has a rather limited dynamic range. To overcome the problem with dynamic range, measurements at low temperature are taken with a perforated anodized aluminum screen to prevent detector saturation. At high temperatures (near $T_{\text{max}}$), the screen is removed to obtain the highest possible sensitivity. At intermediate temperatures, optical measurements are taken both with and without the screen. These last set of measurements are used to scale the low temperature measurements to the high temperature measurements.

### 4.3.3 Spectral measurements

The detection of optical power is insufficient to prove the existence of lasing. A spectral measurement is also necessary to show the existence of a narrow optical linewidth. Spectral measurements were performed first using a Magna-IR 850 Fourier Transform Infrared Spectrometer (FTIR), and later using a Nicolet 8700 FTIR when the former broke down.

Spectra were typically taken with 0.125 cm$^{-1}$ resolution, both at low temperatures and at (or near) $T_{\text{max}}$. Due to the heavy absorption of THz radiation by atmospheric water vapor, the FTIR chamber was usually purged with dry nitrogen gas for measurement. Measurement of spectra at high temperatures were especially challenging, due to the extremely low levels of optical power. Obtaining a spectrum at exactly $T_{\text{max}}$ was not always possible due to the weak power coupled with high atmospheric absorption.
Chapter 5

Waveguide optimization

Terahertz quantum cascade lasers provide very limited gain compared to conventional interband quantum well lasers (tens of cm$^{-1}$ as opposed to hundreds and thousands). Therefore, high operating temperature requires not only the best gain media, but the lowest loss waveguides. This chapter details work done on the optimization of waveguides.

Time domain spectroscopy (TDS) measurements by Burghoff et al. \cite{76} of a scattering assisted gain medium found gain to decrease with temperature at a rate of about 0.1 cm$^{-1}$/K (figure 5-1). Although doubtlessly this rate may change between different laser designs, it does suggest that even one cm$^{-1}$ of loss can substantially impact $T_{\text{max}}$; 0.1 cm$^{-1}$/K implies an improvement of 10 K/cm$^{-1}$ of loss reduction.

5.1 Theory

Consider the waveguide geometry of figure 5-2. Consider the case of very wide laser ridges, such that modal variation in the transverse dimension $y$ may be neglected. Intersubband gain is only available for modes with electric field components parallel to the growth direction, thus QCLs only support tranverse magnetic (TM modes). Therefore, the (complex) magnetic field may be written as

$$\mathbf{H} (\mathbf{r}, t) = \hat{y} H_y (z) \ e^{i(\beta z - \omega t)}$$

(5.1)
Figure 5-1: Gain versus temperature at two different frequencies for a scattering assisted THz QCL, measured using time domain spectroscopy. Reproduced from ref. [76].

Figure 5-2: Typical double-metal waveguide geometry for THz QCLs, where the MBE growth direction is in $z$, and the mode propagation direction is in $x$. Layer thicknesses are not shown to scale.
where $\beta$ is the modal propagation constant and $\omega$ is the lasing frequency. Inserting this result into Maxwell’s equations, we derive the following wave equation for the magnetic $H$-field

$$\frac{\partial}{\partial z} \frac{\varepsilon_0}{\varepsilon} \frac{\partial}{\partial z} H_y + \omega^2 \mu_0 \varepsilon_0 H_y = \beta^2 \frac{\varepsilon_0}{\varepsilon} H_y$$

(5.2)

where $\varepsilon$ is the position dependent permittivity. At the interface between different materials, the continuity of $\frac{\varepsilon_0}{\varepsilon} \frac{\partial}{\partial z} H_y$ is required. This equation is a generalized eigenvalue problem, which can be solved numerically using spectral element methods or similar discretization schemes (see section 2.1).

There are two common waveguides for THz QCLs: semi-insulating surface plasmon (SISP) waveguides, \[77\] and double-metal (MM) waveguides. The latter is unambiguously superior for high temperature performance, \[73\] so is hence the exclusive focus of this thesis.

### 5.2 Material parameters

To solve equation 5.2, the permittivities of every layer in the waveguide structure are approximated using the Drude model.

$$\varepsilon(\omega) = \varepsilon_\infty \left(1 + \frac{i \omega_p^2 \tau^2}{\omega \tau (1 - i \omega \tau)}\right)$$

(5.3)

For metals, $\varepsilon_\infty \approx 1$, and other parameters are taken from from the work of Ordal et al. \[78\][80] Ordal only reports the material permittivity for bulk materials at room temperature, so the applicability of these parameters to thin films under cryogenic conditions is uncertain. However, THz material data is sorely difficult to find, so these parameters are used for a lack of better options. Parameter values for metals used in this thesis are given in table 5.1.

For the $n^+$ contact layers, the parameters taken are $\varepsilon_\infty \approx 10.81 \varepsilon_0$, $\tau = 0.1$ ps. The plasma frequency is calculated using $\omega_p^2 = \frac{ne^2}{m_\epsilon \varepsilon_\infty}$, where the electron concentration $n$ is taken to approximately equal to the doping, and the effective mass is $m_\epsilon = 0.067m_0$. 

135
For GaAs, this yields the numerical result \( \omega_p/2\pi \approx (0.33 \text{ THz}) \sqrt{\frac{n}{10^{16} \text{ cm}^{-3}}} \).

The active region is treated as a “cold-cavity” material, taken to be a lossless dielectric with permittivity \( \varepsilon/\varepsilon_0 = 12.96 \). This is reasonable to the extent that the real index is typically much larger than the imaginary index. Further justification for neglecting imaginary components of \( \varepsilon \) are given in the next section.

5.3 **Active region versus contact losses**

In principle, loss may come from both the active region bulk and the waveguiding metallic contacts. This section addresses the contribution of each component, and justifies the simulation of the active region as a cold cavity (see section 5.2).

5.3.1 **Free carrier absorption**

Traditionally, the QCL active region is assumed to possess free carrier absorption (FCA). This absorption is treated using the Drude model with a plasma frequency calculated using the average QCL doping. However, theoretical results by Carosella et al. [81] and Wacker et al. [82] call this model into question. Loss in the Drude model derives from second-order transitions between free electrons. However, the TM polarization of QCL modes has no electric field in the unconfined directions perpendicular to the growth direction, hence this contribution should be negligible. As such, the only optical absorption possible is that due to intersubband transitions. But such transitions are already modeled in the quantum description of the active region, so this leads to a broader question: does active region “loss” exist distinct from the active region gain spectrum? This thesis argues no. Discussions of active region “loss” amount to an artificial partitioning of the gain spectrum into

\[
g (\omega) = g_{\text{gain}} (\omega) - \alpha_{\text{loss}} (\omega) \quad (5.4)
\]

where \( g_{\text{gain}} \) is the gain from population inverted subband pairs, and \( \alpha_{\text{loss}} \) is loss from noninverted pairs. If \( \alpha_{\text{loss}} \) is included into the waveguide loss calculations, then it must
be excluded from the gain spectrum to avoid double-counting. Given that the loss term $\alpha_{\text{loss}}$ is highly dependent on the bias and design details, writing down a generalized loss term suitable for all gain media is impossible. Therefore, this thesis includes the loss as part of the full spectrum of the gain media, and simulates waveguide loss using cold-cavity simulations.

Although no physical reason exists to include active region loss, a mathematical argument can be made in some circumstances. The reasoning is that the gain spectrum is typically calculated accounting only for first-order (momentum conserving) optical transitions. As such, an active region loss term may still be used to model second-order (momentum displacing) optical transitions. Such second-order transitions do not obey the Drude model, but may nonetheless be considered a form of “free carrier absorption.” However, even this type of second-order transition turns out to be essentially negligible. For example, a theoretical study by ref. [81] found LO phonon emission based FCA to be dominant, but this still only yielded around 1-2 cm$^{-1}$ of loss in the 3-4 THz range.

Further arguments against the existence of free-carrier absorption in the QCL active region are presented in section 5.6.

5.3.2 Comments on Time Domain Spectroscopy (TDS)

MM waveguide losses have been measured by Burghoff et al. [83] and Martl et al. [84] using THz time domain spectroscopy (TDS). Burghoff reports laser power gain to clamp at around 18 cm$^{-1}$ (comments on Martl’s results to follow shortly).

One weakness of the THz TDS technique is that it uses the unbiased QCL as a spectroscopic reference for the biased QCL. The gain is given by the power spectrum of a THz pulse after a single pass through the waveguide when the QCL is biased, normalized against an identically produced pulse passing through the unbiased QCL. In other words, the loss measured is

$$\alpha_{\text{TDS}} (\omega) = \alpha_{\text{on}} (\omega) - \alpha_{\text{off}} (\omega) \quad (5.5)$$
When lasing occurs, $\alpha_{\text{on}}(\omega)$ should be exactly zero at the lasing frequency, and then $\alpha_{\text{TDS}}(\omega) = -\alpha_{\text{off}}(\omega)$. Ideally, the gain medium would be lossless when unbiased, in which case $\alpha_{\text{off}}(\omega) = \alpha_{\text{wg}}(\omega)$. In practice, the unbiased gain medium must contribute some loss as well, so the overall loss can be broken down into two terms, namely,

$$\alpha_{\text{off}}(\omega) = \alpha_{\text{wg}}(\omega) + \alpha_{\text{gm,off}}(\omega) \quad (5.6)$$

where $\alpha_{\text{wg}}(\omega)$ is the loss due to the waveguide, and $\alpha_{\text{gm,off}}(\omega)$ is the absorption of the unbiased gain medium. Thus, $\alpha_{\text{TDS}}(\omega)$ is fundamentally contaminated by the off-state state absorption, and must be regarded as an upper-bound on the true waveguide loss $\alpha_{\text{wg}}(\omega)$.

That said, there is reason to believe that the data contamination by $\alpha_{\text{gm,off}}(\omega)$ is small, at least at low temperatures for some lasing frequencies spectrally distant from low bias absorption features. This originates from comparison of the data of Burghoff \[83\] and Martl \[84\]. As previously mentioned, Burghoff measures gain clamping at 18 cm$^{-1}$ in 10 $\mu$m tall QCL laser bars. Martl measures gain clamping at $\sim$4-5 cm$^{-1}$ electric field gain, which is equivalent to $\sim$10 cm$^{-1}$ power gain, in 16 $\mu$m tall QCL laser bars. Thus, the measurements of Burghoff and Martl scale roughly as the ratio of the QCL active region thickness ($\frac{18 \text{ cm}^{-1}}{10 \text{ cm}^{-1}} = 1.8$, $\frac{16 \text{ nm}}{10 \text{ nm}} = 1.6$). Assuming that both measurements are reliable, this suggest that waveguide losses scale as as the inverse of the active region thickness, or

$$\alpha_{\text{TDS}} \propto \frac{1}{h_{\text{QCL}}} \quad (5.7)$$

This scaling is theoretically expected for losses due to the contact layers and metals. This is consistent with the absence of free carrier effects (section 5.3.1), or active region losses in general.

Equation 5.7 suggests that TDS is an accurate measure of waveguide loss, and that this loss is dominated by contacts layers and metals. This should be at least true for the unbiased QCL. Unfortunately, THz TDS cannot directly answer questions regarding the on-state losses. If free-carrier loss in the QCL active region becomes
significant when the QCL is biased, this would imply a large leakage of electrons into unconfined continuum states. Evidence for significant leakage of carriers to the continuum remains scant at the moment (see chapter 3).

Ongoing work in the THz-MMW group at MIT is looking at the possibility of obtaining absolute $\alpha_{wg}(\omega)$ measurements using self-referencing by the examination of multi-pass THz pulses.

### 5.3.3 Comments on cut-back

Readers from a diode laser background may wonder if a loss measurement can be made through the method of cut-back. Briefly, this involves measuring the differential external quantum efficiency $\eta_d$ as a function of laser cavity length $L$, and fitting it to the expression

$$\frac{1}{\eta_d} = \frac{1}{\eta_i} \left(1 + \frac{\alpha_w}{\ln (1/R)} L\right)$$

(5.8)

from which the internal quantum efficiency $\eta_i$ and waveguide loss $\alpha_w$ can be extracted as a fitting parameters, provided the mirror reflectivity $R$ is known.

This is a well established technique in diode laser engineering, however its application to MM THz waveguides is difficult. The most prominent difficulty is that power measurements of THz lasers are notoriously unreliable. The beam pattern emitted from Fabry-Perot MM waveguides is extremely poor (practically isotropic, even), so collection efficiency is rarely certain. This uncertainty carries over to measurement of $\eta_d$. Furthermore, calculating the mirror reflectivity $R$ is very much a non trivial problem in MM waveguides, as it is dominated by plasmonic rather than Fresnel reflection. Finally, $R$ is much higher in THz waveguides than it is in diode lasers (ranging from 0.7 to 0.9), so the technique is less sensitive than it is for diode lasers.

In summary, extracting $\alpha_w$ from cut-back does not appear likely to work for THz MM waveguides.
Table 5.1: Resistivities at 20°C for select waveguide metals, as well as Drude parameters.

<table>
<thead>
<tr>
<th>Metal</th>
<th>$\rho$ at 20°C (nΩ·m)</th>
<th>$\omega_p/2\pi$ (PHz)</th>
<th>$\tau$ (fs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver (Ag)</td>
<td>15.87</td>
<td>2.18</td>
<td>36.6</td>
</tr>
<tr>
<td>Copper (Cu)</td>
<td>16.78</td>
<td>1.79</td>
<td>72.5</td>
</tr>
<tr>
<td>Gold (Au)</td>
<td>22.14</td>
<td>2.18</td>
<td>24.7</td>
</tr>
<tr>
<td>Aluminum (Al)</td>
<td>28.2</td>
<td>3.57</td>
<td>8.04</td>
</tr>
<tr>
<td>Tungsten (W)</td>
<td>52.8</td>
<td>1.55</td>
<td>10.90</td>
</tr>
<tr>
<td>Tantalum (Ta)</td>
<td>131</td>
<td>1.98</td>
<td>7.54</td>
</tr>
<tr>
<td>Titanium (Ti)</td>
<td>420</td>
<td>0.627</td>
<td>21.8</td>
</tr>
</tbody>
</table>

5.4 Adhesion layer effects on waveguide loss

Section 5.3 suggests that contacts dominate loss, so the rest of this chapter will be devoted to the optimization of the waveguide.

THz frequencies are low enough to be considered “almost” electronic because $\omega \ll \omega_p$ and $\omega \ll 1/\tau$ in metals. The relevance of this statement is that similar to microwave waveguides, low DC resistivity (high conductivity) is a reasonable heuristic for identifying high performance waveguide materials. The room temperature resistivities of select metals are summarized in table 5.1.

The noble metals gold (Au), copper (Cu), and silver (Ag) are good candidates. However, noble metals cannot be deposited directly onto semiconductor surfaces, due to poor adhesion and the possibility of solid-state diffusion between the semiconductor and the noble metal during high temperature processes. An intermediate metal must be deposited that acts both as a tenaciously attached adhesion layer and as a diffusion barrier. A typical adhesion layer for Au is titanium (Ti), and a typical adhesion layer for Cu is tantalum (Ta). This chapter considers the effect of this intermediate layer on waveguide loss.

5.4.1 Ti/Au versus Ta/Cu waveguides

The motivation for this study comes from ref. [23], in which Belkin et al. demonstrated improved temperature performance in Cu waveguides versus Au waveguides. This
result is intuitive, because Cu is a lower resistivity metal than Au. However, an
experiment by Kumar found that Ta/Au and Ta/Cu waveguides give similar results,
suggesting that Ta, rather than Cu, might be responsible for the improvement. In
this section, both Ti/Au and Ta/Cu waveguides are simulated to examine the effect
of either the Ti or Ta adhesion layer thickness.

The waveguide structure of figure 5-2 was simulated, with different adhesion layer
thicknesses (5 nm, 10 nm, 15 nm, and 20 nm). The results of for Ti/Au and Ta/Au are
shown in figure 5-3. Interestingly, the Ti/Au waveguide shows very weak dependence
on the Ti thickness. In contrast, the Ta/Cu waveguide loss is sensitive to the thickness
of the Ta layer, and waveguides with thick Ta show no significant improvement over
Ti/Au.

The reason for this can be understood by examining figure 5-3. Figure 5-3 shows
that the field decays versus distance penetrating into the adhesion layer, but this effect
is more pronounced for Ta versus Ti. Thus, in Ti/Au waveguides, the underlying Au
layer dominates the waveguide loss for all thicknesses of Ti. In contrast, the Ta layer
more effectively screens the field from underlying copper layer, so changing the Ta
thickness controls whether loss is Ta dominated or Cu dominated. Another way of
describing this behavior is through consideration of the skin-depth, which is larger in
Ti versus Ta due to its lower conductivity. Using data from table 5.1 the skin-depth
at 4 THz for Ti is 163 nm, as opposed to 91 nm for Ta.

Contrary to the results of Kumar, these results do predict an improvement of Cu
waveguides over Au waveguides, but these results only manifest when the adhesion
layer sufficiently thin. Kumar used Ta thickness of 30 nm or more.

5.4.2 Effects of the ohmic contacting layer

The QCL active region is typically sandwiched between two heavily doped contact
layers. These layers are important for forming ohmic contacts with the adjacent
metallic waveguides. Kumar, however, showed that etching away of one of these
layers (specifically, the top contact layer) can improve $T_{\text{max}}$. This results in a
Schottky voltage drop across the top contact, which decreases continuous wave (CW)
Figure 5-3: Waveguide loss versus frequency for (a) Ti/Au waveguides, (b) Ta/Cu waveguides, and (c) Ta/Au waveguides. Figure (d) shows the field profile at the metal-semiconductor junction at 3.5 THz. Au/Cu thickness is essentially infinite.
5.4.3 Waveguide optimization to achieve $T_{\text{max}} \sim 200$ K

Sections 5.4.1 and 5.4.2 suggests that loss can be significantly decreased through a combination of thinning the Ta layer in Ta/Cu waveguides and removing the QCL active region top contact layer. These optimizations were implemented in the processing of a THz QCL gain medium provided by researchers at the National Research Council of Canada and the University of Waterloo.

NRC/U. Waterloo’s own fabrication using Ti/Au waveguides and intact top contact layer resulted in a $T_{\text{max}} = 180$ K. At MIT, Ta/Cu waveguides were used and top contact layer was removed. In comparison to refs. [23][33], the Ta adhesion layer was thinned from 20-30 nm down to 10 nm. This improved the $T_{\text{max}}$ to $T_{\text{max}} = 199.5$ K, which is currently the highest $T_{\text{max}}$ record for THz QCLs. The results are summarized in figure 5-5. The success of this project validates the engineering approach described in Figure 5-4: Effect of the removing the top contact layer on double-metal waveguide loss.

performance. However, the characterization of gain media is typically done under pulsed conditions, where heat generation and removal is not as great a concern.

To examine this effect, a 10 nm Ta/Cu waveguide structure was simulated with one contact layer removed. The results are shown in figure 5-4. Compared to the case with an intact top-contact layer, the loss is reduced by around 6-8% in the 3-4 THz range.
5.4.4 Ti/Au vs Ta/Au waveguides

Although sections 5.4.1 and 5.4.3 do support the use of Cu over Au for high performance waveguides, they also highlight the importance of the adhesion layer. For example, figure 5-3 illustrates that Ti/Au can be comparable, or even superior to Ta/Cu in some circumstances, if the Ta layer is too thick. This begs the question about whether there is any inherent material differences in Ti versus Ta as an adhesion layer.

In ref. [89], 5 nm Ti/Au waveguides were found to possess similar performance to 10 nm Ta/Cu waveguides. Intrigued by this work, Ti and Ta were tested by comparing gain medium FL178S-M3 (wafer number VB0517) fabricated in either 10 nm Ta/Au waveguides or 5 nm Ti/Au waveguides. In both cases, the top contact layer was removed.

The results are shown in figure 5-6, which demonstrates a significant drop in $T_{\text{max}}$ moving from 10 nm Ta/Au to 5 nm Ti/Au. This supports the suggestion of Kumar
Figure 5-6: Comparison of 10 nm Ta/Au and 5 nm Ti/Au waveguides. (a) Optical data for 10 nm Ta/Au device (1.69 mm × 150µm, dry etch), with inset spectra. (b) Optical data for 5 nm Ta/Au device (2.06 mm × 150µm, dry etch), with inset spectra. Absolute power was not measured for the device of figure (b).

that Ta may be superior to Ti, but contradicts the findings of [89] (if Ti/Au is inferior to Ta/Au, then it is surely inferior to Ta/Cu as well). A possible explanation for the discrepancy may be that the fabrication process of this thesis uses metal-to-metal thermocompression at 300°C for wafer bonding, whereas the process of ref. [89] uses the much lower temperature transient liquid phase In-Au bonding (∼ 200°C). Perhaps 5 nm of Ti is a sufficient diffusion barrier for Au at ∼ 200°C, but insufficient at 300°C. If this is true, it would nevertheless point to Ta as being a superior metal for fabrication purposes. If it is false, it suggests that Ta is an inherently superior waveguiding metal compared to Ti. In either case, this study conclusively supports Ta as the preferred material for high $T_{\text{max}}$ waveguides.

5.5 Investigations of alternative waveguide structures

The standard waveguide of this thesis is 10 nm Ta/Au, but the results of section 5.4.3 suggests that large improvements in $T_{\text{max}}$ can result from waveguide optimization. This sections investigates some alternative waveguides.
As mentioned in section 5.4, low DC resistivity is a useful heuristic for high performance waveguide materials. Therefore, if Cu improves upon Au, then Ag might be expected to similarly to improve upon Cu. That said, Ag is only slightly less resistive than copper, so any improvement would likely be small. Interesting, according to materials parameters of ref. [78], Ag actually exhibits inferior THz performance to Cu (see figure 5-7). This contrasts the hypothesis of ref. [90], where Ag is predicted to be superior. Therefore, this thesis resolves the question experimentally, by fabricating THz QCLs in Ag waveguides. As an additional optimization, Tungsten (W) was chosen as adhesion layer for silver. W has a much lower resistivity than both Ti and Ta, and there is literature to support its use as a diffusion barrier for Ag. [91]

Since section 5.4 shows that adhesion layer losses can be significant, there is also a possibility that improved waveguides can be crafted by eliminating the adhesion layer altogether. For this purpose, Aluminum waveguides were considered. Although Al is a more lossy metal than the noble metals, Aluminum (Al) adheres to GaAs without the need for an intermediate adhesion layer. [87]

5.5.1 Experimental design

Gain medium FL178S-M3 (wafer number VB0517) was fabricated clad in three different types of waveguide. The first was a 10 nm Ta/Au waveguide fabricated as a reference sample. The second was a 5 nm W/Ag sample. The third was an asymmetric waveguide, with 10 nm Ta/Au for the bottom metal, and Al as the top metal. In all cases, the top contact layer was removed. Figure 5-7 compares the predicted losses for these metals. Theory predicts that the W/Ag guide should perform the best, and that the Ta/Au and Ta/Au-Al waveguides should perform similarly.

The presence of Ag poses some complication for the fabrication process. Ag is an extremely efficient catalyst for H$_2$O$_2$, thus causes the rapid decomposition of etching solutions such as the citric acid based etchant of section 4.1.2, used for MBE substrate removal. For this reason, exposed Ag must be carefully covered with wax or photoresist during the fabrication process.

The asymmetric structure for the Al waveguide is chosen because of fears of inter-
diffusion between the GaAs and Al during the high temperature thermocompression process. Therefore, the bottom metal remains Ta/Au, while the top metal is replaced with Al.

### 5.5.2 Experimental results

The experimental results are summarized in table 5.2. The results do not align with theoretical predictions.

The low temperature $IV$ transport curves of all three devices are plotted together on figure 5-8. The slightly lower $J_{\text{max}}$ of the W/Ag and Ta/Au-Al waveguides is consistent with increased loss.

### 5.6 Comparison to experiment

The losses measured by TDS are higher than calculated losses by a factor of 2-3. There are at least two possible reasons for the discrepancy. The first, and most
obvious is that the material parameters used might simply be incorrect. Materials in THz are generally not well characterized. Second, real waveguide structures may differ from the ideal waveguide structure due to changes induced during fabrication, such as interfacial reactions or diffusion during high temperature steps.

5.7 Conclusions

Predicting waveguide loss from known literature parameters remains a challenging task, and experimental verification is crucial. But from the existing experimental results of this thesis, the following conclusions can be drawn.

- The adhesion layer can be an important determinant of waveguide loss, and should not be neglected in loss calculations.

- The lowest loss THz waveguide found in this thesis is 10 nm Ta/Cu. The top contact layer of the THz QCL MBE stack can be removed to further reduce
loss.

- For fabrication purposes, however, Cu is a more challenging metal to work with due to its chemical sensitivity compared to Au. For general purpose fabrication, where the ultimate performance is not required, this thesis recommends 10 nm Ta/Au. If a gain medium performs well using Ta/Au waveguides, reprocessing using Ta/Cu can be considered afterwards.

- This thesis recommends Ta/Au waveguides over Ti/Au. The loss appears to be lower, although the reasons are not completely clear. Either Ta is a fundamentally superior THz material compared to Ti, or else the difference lies in the interactions of Ta/Ti with surrounding materials (eg. Au/Cu, GaAs) during fabrication.

- No evidence is found to support the use of Ag waveguides. Both experimental and theory (based on the material parameters of ref. 78), suggests that despite Ag’s superior DC conductivity, it makes for an inferior THz waveguide material.

- No evidence is found to support the use of Al waveguides. Despite obviating the need for an adhesion layer, the intrinsic material properties of Al (or possibly the Al-GaAs interface) are unfavorable for THz waveguiding.

- Examination of literature TDS data suggests that waveguide loss is entirely a contact effect, and does not derive from the active region core. (This does not rule out that waveguide loss might scale with doping.)

Clearly, there is still much room for exploration in the optimization of THz waveguides, particularly in materials exploration. However, wafer fabrication in general (and metal deposition in particular) is very expensive. The cost-benefit ratio of such a course of research would likely be high. Furthermore, the extent to which these results can be reproduced in other clean room facilities is not entirely certain; for example, deposition rates and vacuum levels are not always reported, and there may exist other differences between deposition tool characteristics or handling protocols.

This chapter finishes with two simple suggestions for waveguide improvement.
1. Literature sources support the ability of Ta to act as a diffusion barrier for Cu down to 5 nm, [92] so further loss reduction may be possible using 5 nm Ta/Cu.

2. Iridium (Ir) has a much lower DC conductivity than Ti and Ta (47.1 nΩ·m at 20°C), and is also a suitable diffusion barrier for Cu. [93] There is, unfortunately, no information on its THz material properties. Experimental data on Ir/Cu waveguides for THz QCLs would be interesting. Although not a standard metal in MIT’s cleanroom facilities, the existing evaporator should be able to deposit Ir with the right recipe.
Chapter 6

Effects of diagonality on optical gain

As temperatures rise, a major cause of population inversion loss is the degradation of the upper level lifetime due to thermally activated LO phonon scattering. This is illustrated in figure 6-1. One strategy to improve the upper level lifetime is to decrease the wavefunction overlap between the upper and lower lasing subbands; this strategy of pursuing “diagonal” QCL designs was the focus of my master’s thesis, [30] and has also been pursued by other researchers in QCL research. [94]

The strategy of pursuing highly diagonal QCL designs has largely been a failure. The current record holding QCL design is only marginally diagonal, [11] out-performs the more diagonal previous record holder, [24] and lases only slightly hotter than extremely vertical designs. [23] Further attempts to increase diagonality have in fact reduced maximum lasing temperature. The large expected benefits of diagonality have not manifested.

This chapter develops a model to explain the failure of diagonality, and forwards the hypothesis that increased diagonality must be paired with increased doping to achieve its projected benefits. Some experimental evidence is shown to support this theory.
Figure 6-1: Thermally activated LO phonon scattering. Electrons in the upper laser subband can relax to the lower laser subband upon acquisition of sufficient kinetic energy.

6.1 Trade-off between lifetime and oscillator strength \((f_{ul} \tau)\)

The central point of this chapter is that when lifetime is improved by decreasing wavefunction overlap (increasing diagonality), \textit{the gain per unit of population inversion is also reduced.} This is explained as follows.

Let the lasing transition occur between an upper and lower lasing subband \(u\) and \(l\) respectively. Gain is proportional to

\[
g_{ul} \propto \frac{f_{ul} \Delta n_{ul}}{\Gamma_{ul}} \tag{6.1}\]

where \(f_{ul}\) is the transition oscillator strength, \(\Delta n_{ul}\) is the population inversion and \(\Gamma_{ul}\) is the transition linewidth. The oscillator strength is additionally a common metric of the diagonality of a transition, with lower oscillator strengths corresponding to greater diagonality. Therefore, the strategy of moving to greater diagonality essentially \textit{trades oscillator strength for population inversion at high temperatures.} The essential failure of previous work (eg. [30]) lies in the fact that only \(\Delta n_{ul}\) was considered, whereas the
quantity of importance is the product $f_{ul} \Delta n_{ul}$.

This trade-off can be described quantitatively. Consider two subbands separated by less than an LO phonon energy (as is always the case for the upper-lower lasing separation in THz QCLs). Exact expressions for average single body scattering rate between two subbands have been given in chapter \[2\]. However, superior analytical understanding may be had from examining the approximate expression

$$ W_{LO} (u \rightarrow l) \approx W_{LO} (u_{k_{\min}} \rightarrow l) \exp \left( -\frac{\hbar \omega_{LO} - E_u + E_l}{k_B T_u} \right) $$

(6.2)

where $N_{LO}$ is the LO phonon occupancy, and $W_{LO} (u_{k_{\min}} \rightarrow l) = \frac{1}{\tau_{ul}^0}$ is the scattering rate from a state of minimum kinetic energy for LO phonon emission to occur ($\frac{\hbar k^2_{\min}}{2m_e} = \hbar \omega_{LO} - E_u + E_l$; see figure \[6-1\]). The lifetime $\tau_{ul}^0$ is independent of electronic distribution and can be calculated using Fermi’s golden rule.

Empirically, the product $f_{ul} \tau_{ul}^0$ is observed to be roughly a constant, perhaps increasingly slightly as $f_{ul}$ decreases. The mathematical origins of this relationship are not entirely clear, but may be related to presence of the Fourier-transform form factor in the LO phonon scattering matrix element

$$ |A_{ul} (q_z)|^2 = \left| \int dz \psi_u (z) \psi_l (z) e^{i q_z z} \right|^2 $$

(6.3)

where $\psi_{u,l}$ are the upper and lower wavefunctions. Since LO phonon scattering is dominated by transitions with small momentum transfer, one hypothesis is that the contributions of the form factor are dominated by small $q_z$. The use of a Taylor series expansion for the exponential yields

$$ |A_{ul}|^2 \approx q_z^2 \left| \int dz \psi_u (z) z \psi_l (z) \right|^2 = q_z^2 |z_{ul}|^2 $$

(6.4)

which is proportional to the oscillator strength through the dipole moment term.
More generally, all scattering mechanisms, not just LO phonon scattering, should
decrease at the expense of oscillator strength. The essential difficulty of QCL design
is that minimizing scattering matrix elements is challenging without also minimizing
optical matrix elements. Thus, a generalized hypothesis is that $f_{ul}\tau_{ul}$ is roughly a
constant at a given temperature. When LO phonon scattering is dominant, the
product $f_{ul}\tau_{0ul}$ is a constant typically lying in the range of 0.20 ps to 0.25 ps, with a
tendency towards higher values at greater diagonalities (lower $f_{ul}$). This thesis will
hereafter refer to this hypothesis as the “$f_{ul}\tau$ tradeoff.”

6.2 Modeling of 3-well quantum cascade lasers

The 3-well resonant-phonon THz QCL employing 4 subbands is the state-of-the-art
laser architecture for high temperature performance. This system can be modeled
approximately using the effective rates equation approach introduced in chapter 2.
The upper-to-lower scattering rate can be expressed using equation (6.2) to yield an
essentially analytical model of QCL transport.

For the particular model of this section, the relevant transport mechanisms are
indicated in figure 6-2. The parameters associated with this model are given in table
6.1. Solving the rate equations and specifying a particular gain threshold (waveguide
loss) therefore allows us to analyze the maximum lasing temperature as a function of
oscillator strength.

In figure 6-3, the model results are plotted together with a collation of experimen-
tal results representing the best 3-well lasers from the author’s master’s thesis,
as well in the literature. The model produces an good fit to the experimental results
using the stated parameters.

Although the model cannot be considered predictive due to the uncertainty in
many parameters, figure 6-3 nonetheless supports the correctness of its essential
physics. The trade-off between oscillator strength and upper level lifetime poses
a fundamental constraint on the maximum lasing temperature. Diagonality only
marginally improves the maximum lasing temperature, and extreme diagonalities
Figure 6-2: Schematic of analytical model for QCL transport for a 3-well, 4-level QCL. Each horizontal line represents a subband with the symbols denoting (u)pper level, (l)ower level, (e)xtractor, and (i)jector. Parameters for the scattering rates \( \tau \) and tunneling rates \( \tau^* \) are given in table 6.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value/Expression</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{\tau_{43}} )</td>
<td>( \frac{1}{\tau_{43}} + \frac{N_{LO} + 1}{\tau_{43}} \exp \left( -\frac{\hbar \omega_{LO} - E_{43}}{k_B T} \right) )</td>
<td>ps(^{-1})</td>
<td>( u )-to-( l ) leakage</td>
</tr>
<tr>
<td>( \frac{1}{\tau_{34}} )</td>
<td>( \frac{1}{\tau_{34}} + \frac{N_{LO} + 1}{\tau_{34}} )</td>
<td>ps(^{-1})</td>
<td>( l )-to-( u ) backfilling</td>
</tr>
<tr>
<td>( \frac{1}{\tau_{34}} )</td>
<td>( \frac{J_{34}}{0.22 \text{ ps}} )</td>
<td>ps(^{-1})</td>
<td>LO sc. prefactor</td>
</tr>
<tr>
<td>( \frac{1}{\tau_{12}} )</td>
<td>( \frac{1}{10 \text{ ps}} )</td>
<td>ps(^{-1})</td>
<td>( u )-to-( l ) elastic sc.</td>
</tr>
<tr>
<td>( E_{43} )</td>
<td>15 meV</td>
<td>( e ) depopulation</td>
<td></td>
</tr>
<tr>
<td>( \frac{1}{\tau_{41}} )</td>
<td>( \frac{N_{LO} + 1}{0.2 \text{ ps}} )</td>
<td>ps(^{-1})</td>
<td>( e ) backfilling</td>
</tr>
<tr>
<td>( \frac{1}{\tau_{34}} )</td>
<td>( \frac{N_{LO} + 1}{10 \text{ ps}} )</td>
<td>ps(^{-1})</td>
<td>( u )-to-( i ) leakage</td>
</tr>
<tr>
<td>( \frac{1}{\tau_{14}} )</td>
<td>( \frac{N_{LO}}{10 \text{ ps}} )</td>
<td>ps(^{-1})</td>
<td>( i )-to-( u ) backfilling</td>
</tr>
<tr>
<td>( \frac{1}{\tau_{14}} )</td>
<td>( 2 \left( \frac{V_{th}}{\hbar} \right)^2 \tau_{14} )</td>
<td>ps(^{-1})</td>
<td>Injection tunneling</td>
</tr>
<tr>
<td>( \Delta \nu_4 (V_{l4}) )</td>
<td>2.67 (1.33) meV</td>
<td></td>
<td>Inj. anticrossing</td>
</tr>
<tr>
<td>( \frac{1}{\tau_{14}} )</td>
<td>( \frac{1}{2 \tau_{12}} + \frac{1}{2 \tau_{34}} + \frac{1}{0.2 \text{ ps}} )</td>
<td>ps(^{-1})</td>
<td>Inj. dephasing</td>
</tr>
<tr>
<td>( \frac{1}{\tau_{32}} )</td>
<td>( 2 \left( \frac{1}{\tau_{32}} \right)^2 \tau_{32} )</td>
<td>ps(^{-1})</td>
<td>Extraction tunneling</td>
</tr>
<tr>
<td>( \Delta \nu_3 (V_{32}) )</td>
<td>5.0 (2.5) meV</td>
<td></td>
<td>Ext. anticrossing</td>
</tr>
<tr>
<td>( g_{th} )</td>
<td>20 cm(^{-1})</td>
<td>Gain threshold</td>
<td></td>
</tr>
<tr>
<td>( n_{2D} )</td>
<td>( 3 \times 10^{-2} ) cm(^{-1})</td>
<td>Sheet doping</td>
<td></td>
</tr>
<tr>
<td>( L_{mod} )</td>
<td>50 nm</td>
<td>QCL period</td>
<td></td>
</tr>
<tr>
<td>( \Delta \nu )</td>
<td>1.5 THz</td>
<td>Gain-bandwidth</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Summary of parameters used in the semi-analytical model of a 4-level THz described in figure 6-2.
Figure 6-3: Comparison of experimental data versus model of section 6.1 on the relationship between $T_{\text{max}}$ and oscillator strength $f_{43}$. Experimental data is gathered from refs. [23,30,94].
(\(f_{ul} < 0.3\)) actually degrades performance. At some diagonality in the vicinity of 0.1 to 0.15, no lasing is possible (at current levels of doping; see next section).

### 6.3 Overcoming \(f_{ul} \tau\): the need for increased doping in THz QCLs

Adjustments to oscillator strengths and upper-level lifetime amounts to tweaking barrier and well widths in laser design. The essential conclusion of section 6.1 is that such thickness adjustments do not affect the \(f_{ul} \tau\) trade-off, and hence alone are unlikely to boost \(T_{\text{max}}\).

To break the \(f_{ul} \tau\) trade-off, one possibility is to increase QCL doping. Under the assumption that impurity scattering and electron-electron scattering is weak, higher doping simply increases the number of available carriers in a system, and so the available gain should scale linearly with doping. Typical average doping in a THz QCL is only around \(6 \times 10^{15} \text{ cm}^{-3}\); this is 3 to 5 times lower than typically doping in MIR QCLs \((2-3 \times 10^{16} \text{ cm}^{-3})\). Doping levels in the mid- \(10^{15} \text{ cm}^{-3}\) range is based on the work of [17] (apparently, the original THz QCL even failed to lase when its doping was doubled [95]).

Figure 6-3 uses the model of section to illustrate the effects of increased doping, under the assumption that waveguide losses do not scale with doping (see sections 6.6).

Although figures 6-3 and 6-4 are not presumed to have high predictive power, there is an important qualitative conclusion: not only does doping increase \(T_{\text{max}}\), but the optimal oscillator strength decreases as doping increases. Therefore, the central hypothesis is that if the \(f_{ul} \tau\) trade-off is the main limiting factor to high-temperature performance, then improved temperature performance can be attained through THz QCLs which are both highly doped and highly diagonal. Both doping and diagonality are important for success, but prior work concentrated solely on the latter.

The idea of changing the doping in THz QCLs is not new. Refs. [96] and [97]...
studied doping in 4-well, resonant-phonon type THz QCLs, with mixed results; the former concluded the presence of a optimal doping, whereas the latter essentially concluded that doping had negligible effect on $T_{\text{max}}$. What distinguishes this thesis from prior work, however, is that previous studies were performed on very vertical active regions. In vertical active regions, the LO-phonon scattering reduction of the upper level lifetime a much stronger limiting factor.

6.4 First studying of doping: NRC-V812 vs. NRC-V812-M1

The current record-holding device of ref. [1] is one in series of five 3-well devices, designed and grown by collaborators at the National Research Council of Canada, and fabricated and measured by the author at MIT. Another one of these devices, V812, achieved a $T_{\text{max}}$ of 199 K, nearly identical to the record-holding V775 device of ref. [1]. However, V812 possessed a much lower oscillator strength, $f_{43} \sim 0.34$ (as opposed to $f_{43} \sim 0.58$ for V775). Because of this combination of high temperature
performance and high diagonality, V812 was selected as a starting point for testing the doping hypothesis.

V812 was regrown by Dr. John Reno as NRC-V812 and NRC-V812-M1 (wafer numbers VB0521 and VB0523 respectively), with the difference being that doping is doubled in NRC-V812-M1. The bandstructure and band parameters of these designs is shown in figure 6-5.

A brief aside on the calculation of the $f \tau$ product for resonant phonon (RP) designs is in order at this point. RP designs typically have two lower laser levels; these two levels correspond to the anticrossing between a pair of localized subbands, which is the basis assumed by the model of section 6.1. The anticrossed states are expected to “share” the scattering properties of the corresponding localized states. Therefore, to approximate the properties of the localized states, the $f \tau$ is calculated as

$$f \tau = (f_4 + f_2) \left( \tau_4^{0} + \tau_2^{0} \right)$$

Both samples were proccessed into 10 nm Ta/Au waveguides, with dry-etched mesas. The experimental results are shown in figures 6-6 and 6-7. Unfortunately, VB0521 did not perform as well as the original V812 sample from NSERC. A likely possibility...
for this is shown in figure 6-6, which compares the IV transport characteristics of VB0521 and V812. Unlike V812, VB0521 exhibits an negative differential resistance (electrical instability) prior to the attainment of threshold.

In comparison to NRC-V812, however, NRC-V812-M1 achieved a very similar $T_{\text{max}}$ of 196 K. The slightly lower value may be due to the use of Ta/Au instead of Ta/Cu waveguides. Although NRC-V812-M1 did surpass NRC-V812 significantly in $T_{\text{max}}$, the improvement is not as large as expected from figure 6-4. Furthermore, there is no clarity on whether the improvement is due to the greater electrical stability of NRC-V812-M1 or if it is due to the increased doping.

### 6.5 Second study of doping: OWI210H-M series

Because the electrical instabilities of V812 prevented useful conclusions from being drawn in section 6.4, a more electrically stable reference design, OWI210H, was chosen for a second investigation. Figure 6-8 shows the band diagram and experimental data of the original OWI210H (VB0373). As seen in figure 6-8, the design exhibits clear positive differential resistance up to design bias.

Four wafers of OWI210H were grown, designed with different doping concentrations. The designs (wafer numbers) are OWI210H-M4 (VB0610), OWI210H-M5 (VB0612), and OWI210H-M6 (VB0609). Aside from doping concentration, these four wafers also employed bulk doping instead of $\delta$-doping: like in NRC-V812 and NRC-V812-M1, doping was deployed in the middle 4.8 nm of the widest well. This switch was motivated by concerns that at the higher doping concentrations of this study, Si dopants might start to autocompensate under $\delta$-doping conditions.

Nominally, OWI210H-M3 through -M6 were designed for 1-fold, 2-fold, 3-fold and 4-fold the original doping of OWI210H (doping layer doped to 6.2e16, 1.2e17, and 1.9e17 cm$^{-3}$ respectively). In practice however, secondary ion mass spectroscopy (SIMS) analysis performed afterwards revealed that the doping for OWI210H-M3 (VB0605) is roughly twice the designed value (I thank Dr. Asaf Albo for catching this mistake). Consistent with electrical measurements (see section 6-14), the doping
Figure 6-6: Experimental data for NRC-V812 (VB0523). Mesa is dry-etched, 10 nm Ta/Au waveguide, 1.67 mm × 150 µm. (a) Power-Current (LI) (b) Current-Voltage (IV) (c) Spectra (d) IV comparison to V812. NRC-V812 is shifted by -0.8V.
Figure 6-7: Experimental data for NRC-V812-M1 (VB0523). Mesa is dry-etched, 10 nm Ta/Au waveguide, 1.80 mm \times 150 \mu m. (a) Power-Current (LI) (b) Current-Voltage (IV) (c) Spectra. Highest temperature spectra in (c) at 196 K was taken from a smaller device (1.61 mm \times 80 \mu m) due to excessively large current requirements at high temperature in the larger device.
Figure 6-8: (a) Band diagram of OWI210H (VB0373). Layer widths in nm are marked beneath each layer, and dashed line indicate the doping location within the QCL period. (b) Experimental data on OWI210H, reproduced from [30].
Figure 6-9: Secondary Ion Mass Spectroscopy (SIMS) depth profiling of Al, Ga, and Si in VB0605, provided by Evans Analytic Group (EAG), May 16, 2014. Si concentration is determined by comparison to an ion-implanted reference, whose doping is accurate to within ±20%. The SIMS measurement itself has a ±10% tolerance. Therefore, within experimental uncertainty, the measured doping of $6.592 \times 10^{11}$ cm$^{-2}$ over 10 periods is roughly twice of what was designed.

for OWI210H-M4, -M5, and -M6 are believed to be similarly doubled. In effect, this study ended up examining the much larger doping range of 2-fold up to 8-fold standard doping levels.

Figures 6-10, 6-11, 6-12, and 6-13 show the experimental data of each design individually. Figure 6-14 compares the designs across different data sets. As seen in 6-14a, the subthreshold currents for designs M3 through M6 scale roughly with the doping concentration. A better comparison can be made through examination of the IVs scaled by the doping density, shown in figure 6-14b. Figure 6-14b shows that the scaled IVs of OWI210H and OWI210H-M3 agree almost exactly below threshold. For higher dopings, however, the subthreshold tunneling feature (the “knee” in the IV) starts fading.

Most notably, there is a substantial rise in $T_{\text{max}}$ as doping is doubled, rising from 131 K to 177 K. This qualitatively confirms the $f\tau$ hypothesis, and is an encouraging
Figure 6-10: Experimental data for OWI210H-M3 (VB0605). Mesa is dry-etched, 10 nm Ta/Au waveguide, 1.90 mm × 150 µm. (a) Power-Current ($LI$) (b) Current-Voltage ($IV$) (c) Spectra.
Figure 6-11: Experimental data for OWI210H-M4 (VB0610). Mesa is dry-etched, 10 nm Ta/Au waveguide, 2.14 mm \times 150 \mu\text{m}. (a) Power-Current ($LI$) (b) Current-Voltage ($IV$) (c) Spectra.
Figure 6-12: Experimental data for OWI210H-M5 (VB0612). Mesa is dry-etched, 10 nm Ta/Au waveguide, 1.66 mm × 150 µm. (a) Power-Current (LI) (b) Current-Voltage (IV) (c) Spectra.
Figure 6-13: Experimental data for OWI210H-M6 (VB0609). Mesa is dry-etched, 10 nm Ta/Au waveguide, 1.52 mm × 150 µm. (a) Power-Current $(LI)$ (b) Current-Voltage $(IV)$ (c) Spectra.
Figure 6-14: Comparison of experimental data from for OWI210H-M series of devices. (a) $IV$ curves at lower temperature (8 K). (b) Scaled $IV$ curves, where the current density is scaled by the sheet current density (data for the original OWI210H is scaled by half its nominal doping). Voltage shifts applied to each $IV$ to account for differences in Schottky contact voltages are indicated. (c) Threshold current density and maximum lasing temperature vs. doping. (d) Spectra at peak biases at 8 K.
sign for strategy of simultaneously pursuing increased diagonality and doping. However, there are two unexplained points. First, the $T_{\text{max}}$ start deteriorating as doping is increased past 2-fold. Second, the increase in $T_{\text{max}}$ is modest compared to the predictions of figure 6-4. These points are discussed next.

6.5.1 Poisson effects in highly doped THz QCLs

This section addresses the deterioration of $T_{\text{max}}$ as doping is increased past 2-fold standard doping. From figures 6-14c and 6-14d, concurrent with decrease in $T_{\text{max}}$, increasing the doping past $2\times$ also results in a substantial increase in lasing frequency. The increase is not regular to the presence of highly multi-modal lasing found in FP lasers. A slight redshift is also observed going from $1\times$ to $2\times$ doping that be may be due to lasing in different modes. The spectra also seem to broaden somewhat, perhaps under the influence of impurity scattering.

This increase is unexpected, as it suggests an alteration of bandstructure that occurs despite all structures having nominally identical superlattice designs. Some minor differences in growth accuracy exist but these differences are too small to explain the observed blueshift.

Instead these effects are hypothesized to be caused by band-bending due to Poisson effects. Figure 6-15 shows the bandstructure of OWI210H assuming a $55%/5%/5%/45\%$ electron population split between subbands 1 to 4, calculated at the $1'-4$ injection anticrossing. As doping increases, band-bending gets progressively stronger. Given the extreme warping of the bandstructure versus doping, designs OWI210H-M3 through -M6 are unlikely to obey the same transport dynamics. Thus the assumption of a single electron distribution ($55%/5%/5%/45\%$ in this case) for all designs is highly unrealistic. This assumption is made only for simplicity in demonstrating the effects of band-bending.

Two qualitative features can be observed. First, the lasing energy $E_{43}$ increases as

---

1X-ray diffraction data taken at Sandia National Laboratories suggests OWI210H (VB0373) was grown 2.30% low, M3 (VB0605) was grown 2.06% low, M4 (VB0609) was grown 0.09% low, M5 (VB0612) was grown 0.2% high, and M6 (VB0609) was grown 0.23% high.
the doping increases. This is largely due to the increase in subband energy $E_4$ caused by electron charging, and explains the experimentally observed increase in lasing frequency. Second, the band distortion causes the lower laser level and extractor (subbands 2 and 3) to become increasingly misaligned. This subband misalignment is proposed to be the source of the decreasing $T_{\text{max}}$ with increasing doping; heavy doping may have additional adverse effects, such as decreasing the specularity of resonant tunneling through an increase in dephasing.

In standard THz QCL designs, particularly RP THz QCL designs, doping tends to be so low that Poisson effects are largely negligible. [73] This is illustrated theoretically by closeness of figures (6-15a) and (6-15b). Poisson effects have only been relevant for the design of certain long module QCL structures, such as bound-to-continuum designs. In such QCLs, electrons can be more spatially removed from the donor atoms, thus increasing the likelihood of band-bending. However, if the path to higher temperature performance lies in increasing doping, then the present results show that accounting for Poisson effects will be a crucial part of design.

This poses important challenges. In previous work, weak Poisson effects meant that bandstructure could be calculated independent of electron transport, even though transport is strongly dependent on band structure. The addition of strong Poisson effects make bandstructure strongly dependent on electron transport as well, so these two sets of physics must be considered self-consistently.

### 6.5.2 Comparison to simulation

As mentioned previously, the effective rate equations (ERE) simulator of chapter 3 was developed in large part for the accurate treatment of Poisson effects in highly doped designs. In this section, the OWI210H-M series of devices are analyzed using the developed model. The computational results are compared to the experimental results.

Figures (6-16a) and (6-16b) show the calculated IVs and LIs for the OWI210H-M devices. The ERE calculated IVs of figure (6-16a) compare reasonably well with experiment for OWI210H, OWI210H-M3, and OWI210H-M4 (1-fold, 2-fold and 4-
Figure 6-15: Bandstructure of OWI210H-M series devices at injection resonance (1'-4 subband alignment). Layer widths are marked underneath each layer. (a) OWI210H (VB0373) with no band-bending. (b) OWI210H with band-bending. (c) OWI210H-M3. (d) OWI210H-M4 (e) OWI210H-M6. (f) OWI210H-M7. Even in (b), some misalignment occurs between subbands 2 and 3; the problem worsens at higher doping levels.
fold doping respectively), but badly underestimates the currents for OWI210H-M5 and OWI210H-M6.

Due to the many inaccuracies of power measurements, it is difficult to quantitatively compare the calculated \( LI \) curves (figure 6-16b) and the experimental results. However, figure 6-16b qualitatively reproduces the observed experimental trend for the given QCL structure: doping improves \( T_{\text{max}} \) up to a point, after which performance degrades. Unfortunately, quantitative agreement is not produced, as improvement is predicted up to 4-fold doping, where experimentally \( T_{\text{max}} \) degrades after 2-fold doping.

Interestingly, the ERE simulation produces very pessimistic results for the original OWI210H. Lasing is barely expected in this device, according to simulation, where OWI210H was a fairly robust laser in experiment. This discrepancy is under investigation, but one suspected reason is the unusual extraction alignment of OWI210H. Examining figure 6-15a, the lower laser level is designed to be slightly misaligned with the extractor subband at design bias. This choice was implemented by Kumar as means to “extend” the dynamic range of the laser, \([33]\) which is the difference between the maximum current density and the threshold current density \((J_{\text{max}} - J_{\text{th}})\).

In simulation however, this leads to a situation where subbands 2 and 3 are relatively localized, without any means of effective transport between them. Resonant tunneling is not modeled intramodule, and the very small energy difference between them inhibits efficient scattering except by the relatively slow impurity and e-e interactions. This leads to a severe bottleneck of electrons in the lower laser level (subband 3).

Incidentally, the dynamic range \( J_{\text{max}} - J_{\text{th}} \) is a wholly useless quantity. The QCL community has historically considered it very important, with papers “explaining” the poor or good performance of lasers based on their dynamic range (for example, even one of the papers produced by this thesis, \([98]\)). This is misguided. The current \( T_{\text{max}} \) record holder does not have a particularly impressive dynamic range \([1]\), and there are terrible lasers with outstanding dynamics ranges (see for example OWI223R in \([8.1.1]\) and ref. \([99]\)). Not once during the course of this thesis did a good design decision ever arise from consideration of dynamic range.

Although the importance of Poisson effects are demonstrated, the ERE model of
Figure 6-16: Computed (a) $IV$s and (b) $LI$s at a lattice temperature of 8 K for OWI210H-M series of devices. The clamping of $J_{\text{max}}$ above 4x is likely due to the bottleneck of transport from subband 3 to subband 2.

Chapter 3 still clearly needs improvement.

### 6.5.3 High doping and superdiagonality: OWI205H-M7

As part of the same generation of devices as the OWI210H-M series, OWI205H-M7 was designed to embody both extreme diagonality ($f_{43} = 0.13$) and high doping. The results are included here for completeness, although they do not add substantially to the discussion of section 6.5.1.

OWI205H-M7 is OWI210H with a radiative barrier thickened by two monolayers (0.565 nm), and 3-fold standard doping (6-fold after accounting for error in growth). The band diagram both with and without Poisson effects (again assuming a 55%/5%/5%/45 split) is shown in figure 6-17. The experimental results are shown in figure 6-18. The lasing frequency is higher than expected, consistent with results from the highly doped OWI210H-M devices. The $T_{\text{max}}$ of 139 K is unimpressive in itself, but is an interesting result given OWI205H-M7’s extremely low oscillator strength. In figure 6-18a, there are some irregularities in the $LI$ characteristics. For
Figure 6-17: Band diagram of OWI205H-M7. Layer thicknesses in nm are marked between layers. (a) Without band-bending. (b) With band-bending, assuming a 55%/5%/5%/45 split between the first four subbands of the available electrons.
example, at 40 K, power at threshold lases rises steeply initially, but drops abruptly at around 800 A/cm². This is likely due to mode hopping.

OWI205H-M7 is believed to suffer from the same band bending effects as preceding OWI210H-M devices.

### 6.6 Other effects of doping

This section addresses the smaller than expected increase in $T_{\text{max}}$ moving from 1-fold to 2-fold standard folding (figure 6-14). Figure 6-14b shows that the subthreshold transport of OWI210H and OWI210H-M3 are largely identical. Although perhaps not conclusive, this suggests that Poisson effects are yet unimportant when doping is doubled. As such, the smaller than expected increase in $T_{\text{max}}$ is unlikely to be caused by band distortion. This section speculates on other possibilities.

One obvious effect of increased doping is an increase in elastic scattering. Work done on asymmetric dopant diffusion in structurally symmetry active regions in ref. [100] suggests that impurity scattering is a relevant scattering mechanism in AlGaAs THz QCLs, at least when the dopants lie in the active region. Further experiments may be necessary to elucidate the effects of heavy doping on transport and linewidths in THz QCLs.

A second possibility lies in the effects of doping on waveguide loss. Section 6.3 presented doping as a method to defeat the $f\tau$ tradeoff, but it may also be regarded as simply scaling up the QCL gain to defeat waveguide losses. Loosely speaking, the peak gain of a QCL may be written in the form

$$g_{\text{peak}} = C \times n$$  \hspace{1cm} (6.6)

where $n$ is the free electron density (which is proportional to the doping), and $C$ is a constant representing the transport properties of the superlattice (scattering and tunneling). The lasing condition is given by

$$g_{\text{peak}} = C \times n = \alpha$$  \hspace{1cm} (6.7)
Figure 6-18: Experimental data for OWI205H-M7 (VB0611). Mesa is dry-etched, 10 nm Ta/Au waveguide, 2.23 mm × 150 µm. (a) Power-Current ($LI$) (b) Current-Voltage ($IV$) (c) Spectra.
where $\alpha$ is the waveguide loss. If $\alpha$ is constant, increasing doping amounts to increasing $n$ in order to enable lasing at smaller $C$; even unfavorable transport (small $C$) can be overcome by sufficiently large $n$. However, if $\alpha$ increases with doping, then the benefits of increasing $n$ are reduced. In the worst case, total waveguide losses are directly proportional to doping, and increasing $n$ does not help at all.

The success of this chapter suggests that the worst case scenario cannot be true. On the other hand, some increase in loss with doping cannot be entirely ruled out, and may account for the less than expected improvement. A more detailed discussion of waveguide loss is found in chapter 5, in which waveguide engineering is discussed.

### 6.7 Elimination of Poisson effects using two-well designs

Section 6.5.1 highlights the detrimental effects of band-bending. The problem is that a standard resonant-phonon THz QCL is designed to have two resonant-tunneling transitions that align at the same bias. In heavily doped QCLs, band-bending makes it difficult to design this simultaneous alignment correctly without a detailed knowledge of the electron distribution.

However, two tunneling transitions are not strictly necessary. Both Kumar et al. [99] and Scalari et al. [101] have investigated two-well THz QCLs, in which reso-
Figure 6-20: $f_{ul}\tau$ product vs. oscillator strength for a two-well THz QCL. Held constant are the lasing frequency ($E_{32} = 15$ meV), the depopulation gap ($E_{21} = 36$ meV) and injection anticrossing ($\Delta_{1'3} = 2.5$ meV).

Figures 6-20: $f_{ul}\tau$ product vs. oscillator strength for a two-well THz QCL. Held constant are the lasing frequency ($E_{32} = 15$ meV), the depopulation gap ($E_{21} = 36$ meV) and injection anticrossing ($\Delta_{1'3} = 2.5$ meV).

nant tunneling is only used for injection; the lower laser level is directly depopulated by intrawell LO phonon emission. A sample band diagram is shown in figure 6-19. The development of two-well THz QCLs and similar direct-phonon THz QCLs has largely been abandoned, likely because such designs have performed poorly compared to resonant-phonon designs in terms of $T_{\text{max}}$. The reasons for this are not well understood. But such a design cannot suffer from band-bending induced misalignment, since there is no second resonant tunneling transition. This makes it very attractive for developing highly doped THz QCLs.

The two-well design has the additional benefit that its $f_{ul}\tau$ product is much higher than that of resonant-phonon designs. Figure 6-20 shows $f_{32}\tau_{3}^{0}$ calculated (considering LO phonon emission scattering only) as a function of $f_{32}$ in a series of two-well designs. These two-well designs were algorithmically generated to have $E_{32} = 15$ meV, $E_{21} = 36$ meV, and $\Delta_{1'3} = 2.5$ meV (see table 6.2). As seen in figure 6-20 the $f_{ul}\tau$ product is roughly 0.7 ps to 0.9 ps, compared to 0.2 ps to 0.25 ps for resonant-phonon designs.

Nonetheless, there are also dangers to the two-well design. First, the placement of doping is difficult. In traditional THz QCL design, doping in the active region (wells
Table 6.2: Designs represented in figure 6-20. These are generated to have $E_{32} = 15$ meV, $E_{21} = 36$ meV, and $\Delta_{1/3} = 2.5$ meV. Layer thicknesses are given in nm. Band-bending is neglected.

<table>
<thead>
<tr>
<th>$f_{32}$</th>
<th>$\text{Al}<em>{0.15}\text{Ga}</em>{0.85}\text{As}$</th>
<th>GaAs</th>
<th>$\text{Al}<em>{0.15}\text{Ga}</em>{0.85}\text{As}$</th>
<th>GaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>4.821</td>
<td>7.968</td>
<td>5.453</td>
<td>17.107</td>
</tr>
<tr>
<td>0.121</td>
<td>4.830</td>
<td>7.940</td>
<td>5.204</td>
<td>17.100</td>
</tr>
<tr>
<td>0.142</td>
<td>4.838</td>
<td>7.917</td>
<td>4.994</td>
<td>17.092</td>
</tr>
<tr>
<td>0.163</td>
<td>4.844</td>
<td>7.897</td>
<td>4.812</td>
<td>17.083</td>
</tr>
<tr>
<td>0.183</td>
<td>4.850</td>
<td>7.879</td>
<td>4.652</td>
<td>17.073</td>
</tr>
<tr>
<td>0.204</td>
<td>4.854</td>
<td>7.863</td>
<td>4.507</td>
<td>17.062</td>
</tr>
<tr>
<td>0.225</td>
<td>4.858</td>
<td>7.849</td>
<td>4.376</td>
<td>17.050</td>
</tr>
<tr>
<td>0.246</td>
<td>4.861</td>
<td>7.837</td>
<td>4.257</td>
<td>17.038</td>
</tr>
<tr>
<td>0.267</td>
<td>4.864</td>
<td>7.825</td>
<td>4.147</td>
<td>17.025</td>
</tr>
<tr>
<td>0.287</td>
<td>4.866</td>
<td>7.815</td>
<td>4.044</td>
<td>17.011</td>
</tr>
<tr>
<td>0.308</td>
<td>4.868</td>
<td>7.805</td>
<td>3.948</td>
<td>16.997</td>
</tr>
<tr>
<td>0.329</td>
<td>4.869</td>
<td>7.796</td>
<td>3.858</td>
<td>16.982</td>
</tr>
<tr>
<td>0.350</td>
<td>4.871</td>
<td>7.788</td>
<td>3.773</td>
<td>16.966</td>
</tr>
<tr>
<td>0.371</td>
<td>4.871</td>
<td>7.781</td>
<td>3.692</td>
<td>16.950</td>
</tr>
<tr>
<td>0.392</td>
<td>4.872</td>
<td>7.774</td>
<td>3.617</td>
<td>16.933</td>
</tr>
<tr>
<td>0.412</td>
<td>4.872</td>
<td>7.768</td>
<td>3.544</td>
<td>16.916</td>
</tr>
<tr>
<td>0.433</td>
<td>4.872</td>
<td>7.763</td>
<td>3.475</td>
<td>16.898</td>
</tr>
<tr>
<td>0.454</td>
<td>4.872</td>
<td>7.757</td>
<td>3.408</td>
<td>16.879</td>
</tr>
<tr>
<td>0.475</td>
<td>4.872</td>
<td>7.753</td>
<td>3.344</td>
<td>16.859</td>
</tr>
<tr>
<td>0.496</td>
<td>4.871</td>
<td>7.749</td>
<td>3.282</td>
<td>16.839</td>
</tr>
<tr>
<td>0.517</td>
<td>4.871</td>
<td>7.745</td>
<td>3.223</td>
<td>16.819</td>
</tr>
<tr>
<td>0.537</td>
<td>4.869</td>
<td>7.742</td>
<td>3.166</td>
<td>16.797</td>
</tr>
<tr>
<td>0.558</td>
<td>4.868</td>
<td>7.740</td>
<td>3.111</td>
<td>16.775</td>
</tr>
<tr>
<td>0.579</td>
<td>4.867</td>
<td>7.738</td>
<td>3.059</td>
<td>16.751</td>
</tr>
<tr>
<td>0.600</td>
<td>4.865</td>
<td>7.736</td>
<td>3.007</td>
<td>16.727</td>
</tr>
</tbody>
</table>
involved in the optical transition) is avoided in order to keep the gain bandwidth narrow. This is impossible in a two-well laser, as the entire QCL period is the active region. If linewidth broadens linearly with doping, then any improvement in gain using increased doping will be defeated.

Second, the short module length requires that the laser operate at much increased electric fields compared to resonant-phonon designs. This makes it more susceptible to leakages to continuum. Increasing barrier height to combat this might be possible, but again there is a risk of linewidth broadening (due to interface roughness scattering).

Third, the two-well laser is more susceptible to thermal backfilling (as are all direct phonon lasers in general). The extraction tunneling transition in resonant THz QCLs corresponds to an energetically narrow subband doublet in the eigenbasis. Thus, by a density-of-states argument, the lower laser level has an easier time remaining empty, as any electrons scattering to it also have a good chance of scattering to the other subband in the doublet. In contrast, the two-well laser has only a single lower laser level, which is the sole recipient of all backfilled electrons.

Whether the theoretical immunity of the two-well laser to band-bending will outweigh its hazards still needs experimental resolution. This will be the topic of future investigation.

6.8 Conclusion

This chapter presents the hypothesis that present day THz QCLs are limited by a fundamental trade-off between oscillator strength and upper-level lifetime; the product of oscillator strength and upper level lifetime, $f_{ul} \tau$, is postulated to be roughly a constant at any given temperature. As both $f_{ul}$ and $\tau$ are predominantly functions of the barrier-well structure of the QCL, this hypothesis has the important implication that further improvements in THz QCL technology are largely unattainable by simply manipulating layer thicknesses. This traditional mainstay of THz QCL engineering will likely yield incremental improvements at best.

To break the $f_{ul} \tau$ trade-off, this chapter presents a dual approach in which di-
agonality is used to boost the upper level time at high temperatures and increased doping is used to compensate the loss of gain cross section. This strategy is supported by experimental results in which doubling the doping of a highly diagonal \((f \sim 0.2)\) 3-well resonant phonon THz QCL results in a significantly improved \(T_{\text{max}}\).

However, further increases in doping (4-fold increases and higher) results in deteriorating \(T_{\text{max}}\). This is postulated to be caused by increasingly severe Poisson effects (band-bending) at high doping levels. A experimentally observed blueshift in lasing frequency with increasing doping is cited as evidence to support this theory, in connection to the theoretically expected increase of frequency induced by the electron charging of the upper laser level. Experimental verification of this theory is pending.

An important implication is that if the hypothesis of degradation due to Poisson effects is correct, then detailed knowledge of electron distributions are crucial to designing the bandstructure of future QCLs. This means moving beyond the traditional tools of the THz QCL engineer, in which bandstructure is a function only of bias and layer structure. Solvers are needed to self consistently couple transport, bandstructure, and charge distribution; this is the motivation for the work of chapter 3. Alternatively, two-well THz QCLs are suggested to be theoretically immune to these effects (although they possess other disadvantages).

Finally, limitations imposed on the high doping strategy by the relationship between doping and waveguide loss were discussed. Essentially, the success of this strategy relies upon an expectation that gain increases with doping faster than losses do (if they do at all). Experimental evidence is presented to refute the worst case scenario of losses being directly proportional to doping.

To speculate on future directions for design: my belief is that future design could benefit from focusing on maximizing the gain per unit current density at constant doping, rather than maximizing gain itself. The notion is that if maximum gain can be had at minimum current, then the gain can always be bolstered by increasing the doping.
Chapter 7

High Energy Parasitics and the Continuum

Chapter 6.1 discussed the loss of electrons from the upper laser level to the lower laser level. Another posited source of population degradation is the loss of electrons to higher energy states. These can either be the loss of electrons to higher-lying bound subbands, or the leakage of electrons to continuum states. Both mechanisms act as shunting channels that enable electrons to bypass the upper-lower radiative lasing transition.

Compared to the previously discussed thermally activated LO phonon scattering channel, these mechanisms of population inversion degradation are not as well established. However, at a time when progress in THz QCLs seems seriously stalled, even unlikely mechanisms are perhaps worth exploring. This chapter details such efforts.

I have previously published some of the results of this chapter in \[98\] and \[102\]; material has been paraphrased from these publications.
7.1 Elimination of high energy parasitics through ground state designs

Examining the designs of chapter 6, typical resonant phonon THz QCLs always possess 2-3 high energy subbands that lie 40-50 meV above the upper laser level. If these levels do act as shunt-bypasses for the electrons in the upper laser level, one solution is to increase their energetic separation from the upper-laser level, perhaps to 60 meV or even beyond. Unfortunately, there is little room for adjustment in typical RP THz QCLs. The reason for this is that to accommodate the intrawell LO phonon scattering based depopulation of the lower laser level, the widest well must be approximately 15 nm to 16 nm in width to provide a $\sim 36$ meV spacing between the ground subband and first excited subband. This, together with the lasing frequency, essentially locks all other well widths, and hence also locks the position of the higher energy subbands.

In order to have the freedom to adjust the energetic barrier between the upper laser level and high lying subbands, lasers were designed in which depopulation occurs through inter-well, rather than intra-well LO phonon scattering. Unlike in RP designs, where lower level electrons scatter from an excited subband to the ground subband within the same well, the electrons instead scatter from the ground subband of a narrow well to the ground subband of a wider well. Such a design makes no use of excited subbands, and thus has been dubbed a ground-state THz QCL (GS THz QCL).

This ground-state concept is not new. In fact, because of its simplicity, the first THz emitting structure based on LO-phonon scattering for depopulation employed this scheme, but lasing has never been experimentally realized.

7.1.1 Design of a 3-level ground-state design (OWIGS271)

The simplest ground state design consist of an upper laser subband, a lower laser subband, and an injector to collect electrons from the lower laser level. This is implemented in design OWIGS271, whose band-diagram is shown in fig. 7-1a.
design is qualitatively the same as the one reported in ref. [104]. Even though it has the same number of wells as the 3-well RP designs, [1, 24, 105] there is a qualitative difference. The RP design has two subband levels in the phonon well (widest well) involved in electron transport, with the excited subband aligned with the lower lasing level at the design bias. Also, unlike most RP designs, 30% Al instead of 15% Al barriers are used.

The closest high-energy parasitics is 57 meV away in OWIGS271. For comparison, in the 200 K design of ref. [1], the closest parasitic subband is 39 meV away from the upper laser level. This greater energy barrier should strongly suppress undesired interactions in the latter.

There are several necessary trade-offs in the design. The lack of resonant tunneling in extraction reduces the selectivity of the extraction process. Thus, if the lower level lifetime is kept short (<1ps) by using a thin barrier, then the upper level lifetime can be preserved only by spatially separating the upper and lower lasing subbands; in other words, the radiative transition must be highly diagonal in space (the radiative oscillator strength normalized to the effective mass of GaAs for OWIGS271 is \( f_{32} = 0.20 \)). Accounting only for LO phonon scattering, the calculated upper and lower level lifetimes at high temperatures are 1.07 ps and 0.58 ps. While this depopulation is slow compared to the lifetime of the intrawell LO phonon emission commonly employed in RP THz QCLs, no transit time is needed for tunneling. [56]

7.1.2 Experimental results of the 3-level ground-state design

OWIGS271 was grown in wafer number VA0323 and processed into 10 nm Ta/Cu waveguides with wet-etched mesas. Experimental data is shown in figure [7-2] in which lasing at \( \sim 5.2 \) THz is observed up to a heat sink temperature of 71 K.

Although OWIGS271 has identical bulk doping as published RP designs, its maximum current density is much lower. This is due to the extremely diagonal optical transition, which leads to the upper level lifetime being the bottleneck to transport. Examination of the \( I - V \) characteristics reveals that OWIGS271 also shows a pronounced shoulder feature at biases below the onset of lasing operation, which is likely
Figure 7-1: (a) One module band diagram for OWIGS271. Layers are marked by their thicknesses in nm. The 8.2 nm well is doped to $2.4 \times 10^{16} \text{ cm}^{-3}$. (b) Two module band diagram for OWIGS271, showing parasitic interaction between injector and lower laser level (1'–2).
due to the resonant interaction of the injector subband with the lower lasing level in the next QCL period (see figure 7-2(b)). This is a well known issue in one-well injector RP designs. This early NDR might be one cause of poor temperature performance.

Figure 7-2: Experimental data for OWIGS271 clad in Ta/Cu waveguides with wet-etched mesas.

### 7.1.3 Design of a 4-level ground-state design

In order to reduce the coupling between the injector and lower lasing levels seen in OWIGS271, a second GS design, TWIGS254, was designed that employs a two-well injector. The band diagram of TWIGS254 is shown in figure 7-3. The oscillator strength and frequency ($f_{43} = 0.19$, $E_{43} = 21.5$ meV) are similar to OWIGS254.

### 7.1.4 Experimental results of the 4-level ground-state design

TWIGS254 was grown in wafer number VA0336 and processed similarly to OWIGS271, except that Ta/Au waveguides were employed in place of Ta/Cu, and ridges were de-
Figure 7-3: One module band diagram for TWIGS254. Layers are labeled by their thicknesses in nm. The 3.4 nm barrier is delta-doped in the middle to $2.7 \times 10^{10}$ cm$^{-2}$.

...lunged by dry-etching. Experimental data is shown in figure 7-4. Lasing at $\sim$5.2 THz is observed up to a heatsink temperature of 25 K. While the lower level parasitic is strongly suppressed (the shoulder feature at $\sim$11 V is barely visible) and positive differential resistance is restored through, TWIGS254 performed far worse than OWIGS254 in terms of operating temperature. For reasons unclear (perhaps due to differences in growth and fabrication), the lasing threshold ($\sim$ 400A/cm$^{2}$) is much higher than that for OWIGS254 ($\sim$ 310A/cm$^{2}$).

7.1.5 Comparison of OWIGS271 and TWIGS254, and other discussion

Although lasing action in both designs was achieved, the maximum operating temperatures are rather low. A possible cause of overall poor temperature performance is the unusually large radiative gap chosen for both designs ($\sim$ 20 meV). This choice was motivated by a desire to test the extents of the empirical rule-of-thumb for THz QCLs that $T_{\text{max}} \leq h\nu/k_B$. Kumar et al. in ref. [106] speculate that this result originates from the loss of injection selectivity as the radiative gap shrinks. Conversely,
Figure 7-4: Experimental data for TWIGS254 clad in Ta/Au waveguides with dry-etched is mesas. (a) L-I versus temperature. (b) I-V at 8K. The dotted lines mark the lasing dynamic range in current (compare (a)). The arrow marks the parasitic shoulder attributed to alignment between the injector and lower laser levels.

larger radiative gaps are more susceptible to thermally activated LO phonon scattering, which is postulated to be the dominant cause of temperature degradation in
THz QCLs. \( T_{\text{max}} \leq h\nu/k_B \) has not been previously examined at such high lasing frequencies (>5 THz), and it is possible that LO phonon scattering induced population inversion degradation overwhelms any benefit to be had from improved injection efficiency.

7.1.6 \( f_{ul}\tau \) tradeoff in Ground State Designs

As discussed in chapter, THz QCLs are hypothesized to be limited by a fundamental tradeoff between the oscillator strength of the lasing transition and the upper level lifetime (\( f_{ul}\tau \)). In chapter 6 this was discussed in the context of RP designs, but this should apply to the ground-state designs considered here as well. Following demonstration of lasing in sections 7.1.2 and 7.1.4 designs with even narrower wells and greater energetic barriers were designed, but the resulting designs were observed to have very small \( f_{ul}\tau \) products compared to resonant phonon designs (ie. less than 0.2 ps).

To investigate further the origins of this, the following numerical experiment was performed. Two quantum wells were simulated. The width of well was varied from 17.0 nm down to 6.8 nm, while the width of the second well and thickness of intermediate barrier were adjusted to keep a fixed oscillator strength and energy separation. This emulates the design strategy of moving to narrower wells to improved energetic separation from high lying subbands. The \( f_{ul}\tau \) product was calculated for each well width, using LO phonon scattering. The results are plotted in figure 7-5 plotted against the energy of the lower subband. The conclusion is that even at constant oscillator strength, the \( f_{ul}\tau \) product decreases as the wells are made narrower.

Thus, although the energetic barrier separating higher lying states can be made arbitrarily larger by narrowing the wells, the \( f_{ul}\tau \) tradeoff worsens the more this is done.
Figure 7-5: (a) Schematic of numerical experiment. The energy $E_{21}$ and its associated oscillator strength $f_{21}$ are fixed at 15 meV and 0.20, respectively. (b) Under the conditions of (a), the $f_{ul}\tau$ product for $2 \rightarrow 1$ under variation of $E_1$ through changing the well width $w_1$ (and to a lesser extent $w_2$ and $b$).

7.1.7 Conclusion

While the GS design strategy may be useful for reducing shunt current leakages through high-energy subbands, forcing the radiative transition high into the conduction band results in an increasingly unfavorable trade-off between oscillator strength and lifetime. At present, the former problem is less well established than the latter, so the GS strategy does not seem favorable.

If the $f_{ul}\tau$ trade-off can be fixed through increased doping, as posited in chapter 6 then shunt current channels may become a new limiting factor in THz QCL design. The GS strategy may be worth revisiting at such a time, should it come.

7.2 Suppression of leakages to continuum using tall barrier designs

Continuum states are defined as quantum states with ($z$-directed) energy lying above the barrier heights. As discussed in 2.5.1 above-barrier electrons are not bound in
the classical sense, and thus are free to accelerate in the presence of an electric field. Just like the suppression of interactions with high-energy subbands, interactions of the transport subbands with continuum states can be suppressed by increasing their separation in energy. The design in this case, however is easier, as it corresponds to simply increasing the barrier height.

### 7.2.1 Design of a tall injector THz QCL (NRC-V775C)

In MIR QCLs, extremely tall conduction band barriers are sometimes used to suppress leakages to continuum. Botez et al. have also used variable barrier heights in MIR QCLs to engineer leakage reduction, although their motivation for doing so was to decrease shunting channels caused by high energy parasitic subbands. Inspired by these successes, a GaAs/Al$_{0.15}$Ga$_{0.85}$As THz QCL was designed in which the injector barriers are replaced with pure AlAs.

The idea of employing multiple Al fractions in the same QCL design has been explored in ref. [110], in which Monte-Carlo simulations predict that designs employing multiple Al fractions yield superior gain and temperature performance. Unfortunately, experimental results do not support these predictions. Nevertheless, increasing the design-space of THz QCLs by removing the constraint of using only a single Al fraction can be reasonably hypothesized to yield some improvement in performance.

An important advantage of the present work over ref. [111] is that AlAs is a stoichiometric binary material, and requires no additional calibration of Al flux during growth by molecular beam epitaxy (MBE). Another advantage is that the binary AlAs presumably also reduces alloy scattering compared to the ternary Al$_{0.15}$Ga$_{0.85}$As barriers, although an enhanced interface roughness scattering associated with thin AlAs barriers may negate this perceived advantage.

The structure, named NRC-V775C, is a modification of the $T_{\text{max}} \sim 200$ K design reported in ref. [1]. The injector barrier is replaced with pure AlAs with a thickness chosen to maintain the same injection coupling ($2\hbar\Omega_{1+4} \sim 2.7$ meV), with all other wells and barriers tweaked to maintain roughly the same energy separations and
diagonality. The injector barrier was chosen for replacement with AlAs because it is the thickest barrier. Replacement of all Al$_{0.15}$Ga$_{0.85}$As barriers with pure AlAs is difficult because the small subband separations in THz QCLs would in some cases require sub-monolayer barriers. For comparison, the design of ref. [1] was regrown in the same MBE machine as structure NRC-V775A.

The bandstructures for NRC-V775C and NRC-V775A are shown in figure 7.6, calculated using the methods of chapter and material parameters from ref. [112]. Following the recommendations of ref. [112], discontinuities of 0.149 eV and 1.050 eV are used for GaAs/Al$_{0.15}$Ga$_{0.85}$As and GaAs/AlAs interfaces, respectively. However, the device of ref. [1] was designed using 0.135 eV for GaAs/Al$_{0.15}$Ga$_{0.85}$As, so the barriers of NRC-V775C are somewhat thinner. At the time, the changes in band parameters were felt to be minor, but in retrospect, this study could have been better designed: the band parameters for NRC-V775A should have been recalculated using Vurgaftman’s material parameters prior to the matching between NRC-V775A and NRC-V775C.

7.2.2 Experimental results of tall injector design

NRC-V775A was grown in wafer number VB0486, and NRC-V775C was grown in wafer number VB0487. Both wafers were processed into dry-etched laser bars with 10 nm Ta/Au double metal waveguides. Experimental results are shown in figures 7.7 and 7.8.

In figure 7.7, the increase of the threshold current as temperature rises is clearly slower in NRC-V775C, consistent with a reduction in parasitic leakages. But surprisingly, NRC-V775C did not lase below $\sim$70 K. Despite this, NRC-V775C lased up to 181 K in pulsed mode, whereas NRC-V775A lased up to 175 K. Both structures were inferior to the original $\sim$200 K structure in ref. [1]. NRC-V775C was overgrown by 1.2%, and NRC-V775A was undergrown by 1.8%, so growth differences may be at play. We note in figure 7.7b that the optical output starts decreasing prior to the negative differential resistance feature (NDR) near the design bias. This was not seen in the original structure, and may point to slight problems in the growth. Part of the drop in $T_{\text{max}}$ may also be caused by the use of Ta/Au waveguides instead of
Figure 7-6: Band diagrams for (a) NRC-V775C and (b) NRC-V775A. Layers are marked by their thicknesses in nm, and one module is boxed in each case. In both cases, the middle 5 nm of the widest well is doped to $6 \times 10^{15}$ cm$^{-3}$. 
Ta/Cu. Also the room temperature device resistance is unusually low, suggesting that the top-contact layer may have been incompletely etched away. This last possibility would be consistent with ref. [1], in which different samples with intact doping layers and Ti/Au waveguides lased up to 176 K and 180 K.

Rather than something exotic like DX centers or indirect barrier states associated with the AlAs barriers, we hypothesize that the failure to lase below ~70 K is due to the occurrence of NDR just before the threshold. This is known to harm the performance of THz QCLs, especially three well designs of this type which can suffer from voltage instabilities. Although the extraction couplings for NRC-V775C and NRC-V775A are designed to be identical, NRC-V775C was designed with a slightly taller barrier height for Al$_{0.15}$Ga$_{0.85}$As barriers ($2\hbar\Omega_{23} \sim 4.9$ meV was targeted, but $2\hbar\Omega_{23} \sim 5.5$ meV if calculated using 135 meV barriers). Consequently, its extraction barrier is thinner (3.7 nm instead of 4.1 nm), which results in a stronger interaction between the injector of one QCL module and phonon well excited state of the next (the intermediate resonance described in ref. [114]). That said, it remains unclear why the lasing is so thoroughly eliminated, whereas the devices of ref. [94] lased despite the presence of an early NDR.

The lasing frequency at low temperature is lower than expected from a one-module simulation. One possibility that injector anticrossing is strong enough to enable coherent coupling between the injector subband and the upper laser level, which pushes down the energy of the latter. Another possibility is that interdiffusion is occurring between the injector barrier and its adjacent wells. In ref. [114], interdiffusion lengths of up to 0.8 nm in AlGaAs/GaAs are reported (notably, the authors of ref. [114] are also supplied with MBE growth by Dr. John Reno, so their results are directly applicable to work at MIT). Since the diffusion lengths are comparable to the barrier width, this should result in barrier height lowering. The reason the interdiffusion does not have a significant effect on conventional designs seems to be related to the fact that the barriers in 15% designs are much thicker. I thank Prof. Oana Malis and Prof. Karl Unterrainer for useful comments on the possibility of AlAs/GaAs interdiffusion.

The question, then, is whether this study truly demonstrates suppression of leak-
Figure 7-7: Experimental \textit{LIVT} and spectral data for (a) NRC-V775C \((1.96 \text{ mm} \times 150 \text{ µm} \times 10 \text{ µm} \text{ ridge})\) and (b) NRC-V775A \((1.92 \text{ mm} \times 150 \text{ µm} \times 10 \text{ µm} \text{ ridge})\). In (a), the 71K \textit{IV} is rough due to the intermittent onset of lasing.
Figure 7-8: Comparison of threshold current rise between NRC-V775C and NRC-V775A

age to continuum or not. The increase in $T_{\text{max}}$ is minimal between the two designs. The possibility that the increase is merely experimental variation cannot be dismissed. However, if the signal is real and the improvement is due to reduction of the leakage to continuum, this would suggest leakage to continuum is not a major scattering channel. Is it then possible that the improvement is real, but the cause is not suppression of leakage to continuum? One possibility is that the cause is the slightly thinner barriers improved the coherence of tunneling.

To argue in favor of continuum suppression, figure 7-8 shows that the threshold current density rises less steeply with increasing temperature in NRC-V775C compared to NRC-V775A. However, stable threshold correlates poorly with high $T_{\text{max}}$, even in MIR QCLs. [107] ie. many THz QCLs have a stable threshold and still have a terrible $T_{\text{max}}$.

### 7.2.3 Results for an all tall barrier design

Encouraged by the better performance of NRC-V775C over NRC-V775A, a three-well resonant phonon active region was designed with all AlAs barriers, shown in figure 7-9a. A very diagonal lasing transition is employed, as a vertical transition would
require excessively thin barriers. The structure, OWI230T, was similarly grown and processed (except that ridges were wet etched), but failed to lase. The IVs are presented in figure 7.9b, which shows an NDR feature persisting up to \( \sim 170 \) K.

If this NDR feature is due to the same intermediate resonance, it may be that early NDR prevents lasing at any temperature in this structure. On the other hand, OWI230T may also have failed due to excessive interface roughness scattering associated with the thin AlAs barriers. The importance of interface roughness scattering for THz QCLs has been theoretically predicted \cite{46} and is further supported by recent experimental results on interface roughness scattering in nominally symmetric active regions. \cite{14} That said, vertical correlations between the interfaces in such thin barriers should reduce the rates somewhat, \cite{115} and the golden rule may not be applicable to barriers so thin that they might host voids. \cite{116} This result from a highly diagonal QCL would be consistent with the results of ref. \cite{117}, in which the barrier Al fraction was systematically increased for a vertical QCL design; when increased to \( \text{Al}_{0.45}\text{Ga}_{0.55}\text{As} \), the design similarly failed to lase.

### 7.2.4 Conclusion

Other researchers have previously attempted to improve THz QCL performance by increasing barrier, either as the primary design modification or a secondary modification made to accommodate other design features (eg. taller barriers are often needed to accommodate designs which emit multiple LO-phonon emissions per period). \cite{118} These previous efforts have uniformly met with failure, with the resultant designs nearly always exhibiting lower \( T_{\text{max}} \) compared to reference 15% Al designs. In light of this, the first significance of this work is that it is (one of) \cite{11} the first demonstration(s) of a tall barrier design providing superior performance to a reference 15% design. The posited mechanism for improvement is suppression of leakage to con-

\[ ^{1} \text{I have recently discovered that a 7 K increase in } T_{\text{max}} \text{ was demonstrated by Lin et al. by moving from 15% to 30% Al barriers.} \]
Figure 7-9: OWI230T (a) band diagram and (b) experimental IVs (1.32 mm × 100 µm × 10 µm ridge). Layers in (a) are marked by their thicknesses in nm, and one module is boxed; the middle 4.8 nm of the widest well is doped to $2.5 \times 10^{17}$ cm$^{-3}$. 
tinuum; even if this true, leakage to continuum does not appear to be a major temperature degradation mechanism in THz QCLs.

The second significance is that it suggests the possibility of a laser which works better at higher temperatures than at lower temperatures. If a room temperature THz QCL exists, it may be the case that it works at 300 K but not at 4 K. Such behaviour has been predicted theoretically for a different QCL structure. [120]
Chapter 8

Terahertz QCL Gain Media for Practical Applications

While the pursuit of room temperature Terahertz Quantum Cascade lasers is an important endeavor, for some purposes, cryogenic cooling is acceptable. When this is the case, other design criteria become more important than the optimization of gain or temperature performance alone. This chapter details certain projects undertaken in applied THz QCLs.

8.1 Terahertz QCLs for oxygen line detection

As mentioned in chapter [1], astrophysics is a major application area of terahertz technology, through its applications to spectroscopy. For many years, THz-MMW at MIT was a member of the Galactic/Xtragalactic ULDB Spectroscopic Stratospheric Terahertz Observatory (GUSSTO) project, led by Prof. Chris Walker at the University of Arizona. GUSSTO intended to send a 1-meter telescope equipped with heterodyne receiver arrays at 1.4 to 4.7 THz. The role of THz-MMW in GUSSTO was to provide a 4.7 THz local oscillator for a heterodyne spectroscopy system built at the Technical University of Delft; [12] the relevance of 4.7 THz is that 4.74 THz is a neutral oxygen (OI) emission line that shines particularly bright around massive star formation, and hence is useful for studying interstellar physics. [12]
Unfortunately, the best THz QCL designs almost always lie in the 3-4 THz spectral region, thus GUSSTO called for the development of high performance, high frequency THz QCLs. Sadly, GUSSTO never moved forward past its initial exploration stages, but prior to its premature termination, work was carried out at MIT to create 4.7 THz gain media. These efforts are documented in this section.

8.1.1 4.7 THz QCLs based on 3-well resonant phonon active regions

The 3-well resonant phonon architecture is the current state-of-the-art in THz QCL design, at least for high temperature performance. Therefore, a 4.7 THz 3-well QCL was attempted. To the best of my knowledge, this work represents the first attempt at a 4.7 THz 3-well laser, thus the only prior art available to assist in design is results from other lasers architectures.

One major question was what diagonality would be optimal for such a design. As the laser frequency increases, the upper-lower subband energy difference approaches the LO phonon energy, thus shortening the upper level lifetime. The necessity of higher diagonality is therefore expected. That said, Williams and Kumar designed high frequencies QCLs at $\sim 4.3$ THz \cite{122} and $\sim 4.8$ THz \cite{33} respectively with essentially vertical active regions. Thus, the first design attempt was made only moderately diagonal, with $f \sim 0.4$.

Furthermore, experience indicates that lasing often occurs at lower than expected frequencies, which is attributed to the influence of the injection anticrossing. Therefore, the single-module radiative transition was designed slightly higher than 19.4 meV, which corresponds to 4.7 THz. Finally, consistent with the record holding $\sim 200$ K design of ref. \cite{1}, barriers are kept rather thin; the injection and extraction anticrossings are 2.76 meV and 4.99 meV respectively. Physical justification for the choice of thin barriers (equivalently, large anticrossings) lies in the fact that injection and extraction selectivity are less important for high frequency designs (in one extreme, MIR QCLs can use anticrossings measured in tens of meV).
Figure 8-1: Band diagram of OWI223R. Well and barrier widths in nm are indicated. The doped portion of the design is between the dashed lines.

The band diagram of the final design, OWI223R, is shown in figure 8-1. OWI223R was grown three times, as wafers VB0488 (overgrown by 3.9%), VB0492 (undergrown by 2.9%), and VB0493 (undergrown by 0.9%). The experimental results of the most accurate grown wafer, VB0493, are shown in figure 8-2. The resultant laser is evidently quite poor, having both bad temperature performance \( T_{\text{max}} = 71 \text{ K} \) and overly high frequency \( \sim 5.1 \text{ THz} \). The experimental results of VB0492 are shown in figure 8-3. VB0492 had similar performance, with a slightly improved \( T_{\text{max}} \) of 76 K; interestingly, despite being more under-grown than VB0492, the lasing frequency actually dropped and is very close to 4.7 THz. Finally, the experimental results of VB0488 are shown in figure 8-4. Surprisingly, this severely overgrown wafer had substantially better \( T_{\text{max}} \) than VB0492 and VB0493. VB0488 also lased very close to 4.7 THz. The fact that both the under-grown VB0492 and over-grown VB0488 lased closer to the target frequency than the relatively accurate VB0493 is unintuitive, and highlights the difficulties of predicting laser frequency from design.

OWI223R did not show evidence of early NDR, so the mechanisms of its poor performance are likely due to weaknesses in regular transport. Guided by the superior performance of VB0488 compared to the VB0492 and VB0493, the poor performance of OWI223R was attributed to upper level leakage. VB0488’s overgrowth likely re-
Figure 8-2: Experimental data for OWI223 (VB0493). Device dimensions are 1.96 mm × 150 μm. (a) LI (b) IV (c) Spectra.
Figure 8-3: Experimental data for OWI223 (VB0492). Device dimensions are 1.91 mm × 150 µm. (a) $LI$ (b) $IV$ (c) Spectra.
Figure 8-4: Experimental data for OWI223 (VB0488). Device dimensions are 1.78 mm × 150 µm. (a) LI (b) IV (c) Spectra.
resulted in thicker barriers, which may have improved its performance. Therefore the next iteration of OWI-R was made more diagonal, and the radiative transition was reduced slightly to modify the frequency. The doping was also increased by 50%, in keeping with the $f\tau$ hypothesis of section 6.1.

The band diagram of the modified design, OWI219R-M1 is shown in figure 8-5. Three wafers were grown, VB0524 (over-grown by 0.98%), VB0540 (over-grown by 2.9% high), and VB0541 (over-grown by 1.3%). VB0541 proved a much better laser than the growths of OWI223R, with a $T_{\text{max}}$ of 127 K, and lasing at the desired 4.7 THz (figure 8-8). Similar results are found for VB0524 (figure 8-6) and VB0541 (figure 8-7).

VB0524 was noted, however, to exhibit unusual spikes in power at different temperatures. The behaviour was not entirely reproducible, and appears to be related to unstable lasing in pulsed mode. The detector signal was noted to fluctuate significantly when monitored with the oscilloscope, but these fluctuations are washed out by the influence of the lock-in amplifier during measurement.

Furthermore, at very high biases (past $J_{\text{max}}$ and the principle NDR), all three wafers showed traces of lasing at the extraordinarily high frequency of $\sim 6.2$ THz (figures 8-6c, 8-7c, and 8-8c). As these measurements were completed in pulsed op-
Figure 8-6: Experimental data for OWI219R-M1 (VB0524). Device size is 1.31 mm × 150 µm (a) LI (b) IV (c) Spectra (d) Peak power vs temperature. Lasing was noted to be unstable in this device; perhaps at certain temperatures stabilization occurred, resulting in higher power.
Figure 8-7: Experimental data for OWI219R-M1 (VB0540). Device size is 1.31 mm × 150 μm (a) LI (b) IV (c) Spectra
Figure 8-8: Experimental data for OWI219R-M1 (VB0541). Device size is 1.76 mm × 150 µm. (a) LI (b) IV (c) Spectra
eration, the stability of this high frequency lasing is uncertain; a continuous wave (CW) measurement for confirmation would have been ideal, but this measurement was not completed due to other research priorities taking precedence at the time. However, these results do raise the question of what is the maximum lasing frequency achievable by THz QCL technology. As EM waves cannot propagate in GaAs within the Restrahlen band, one physical limit must be the transverse-optical (TO) phonon frequency of GaAs (~8 THz). Is there, however, other limits in place below the TO frequency?

8.1.2 4.7 THz QCLs based on 4-well resonant phonon designs

Section 8.1.1 essentially details the development of an application specific THz QCL from scratch. However, as mentioned previously, Williams and Kumar have had success in making high frequency THz QCLs based on the original resonant phonon architecture. This section details the parallel development of 4.7 THz QCLs based on the antecedent set by Williams and Kumar.

FL183R was originally designed to be a 4.7 THz laser, but experimentally lased at 4.3 THz (possibly due to the influence of the injector anticrossing). Nevertheless, it has proven to be one of the highest power THz QCLs in existence. Therefore, FL183R was chosen as the starting point for further development.

Because FL183R had not been far off from achieving 4.7 THz lasing, drastic modifications were not attempted. However, data from OWI223R established that high frequency designs require greater diagonality for optimal performance. Therefore, FL182R-M6 and FL181R-M7 were designed as minimal modifications of FL183R. The principle difference is that FL182R-M6 has its radiative barrier thickness increased by 1 ML, and FL181R-M7 has its radiative barrier thickness increased by 2 ML. The wells adjacent to the radiative barrier are adjusted to correct the radiative transition energy. Parallel to the development of OWI219R-M1, the doping in each design was also increased by 50%. The band diagrams are shown in figures and 8-9 and 8-10.

FL182R-M6 was grown as wafer VB0520, and FL181R-M7 was grown as wafer
Figure 8-9: Band diagram of FL182R-M6. Well and barrier widths in nm are indicated. The doped portion of the design is between the dashed lines.

Figure 8-10: Band diagram of FL181R-M7. Well and barrier widths in nm are indicated. The doped portion of the design is between the dashed lines.
VB0519. The experimental results are shown in figures 8-11 and 8-12. VB0520 lased at slightly too low frequencies, but had an excellent $T_{\text{max}}$ of 153 K for a laser of such high frequency. Conversely, VB0518 lased at essentially the correct frequency (perhaps slightly too high), but had a much worst $T_{\text{max}}$ of 127 K.

8.1.3 Conclusion

Strong 4.7 THz gain media based on 3-well and 4-well resonant-phonon QCL architectures were successfully designed and fabricated. That said, lasers at this frequency still lag behind the state-of-the-art 3-4 THz lasers, and work remains to be done (pending the need for such lasers, of course). The $T_{\text{max}}$ of 4.7 THz laser remains rather low, and reducing their current density to minimize power consumption will be beneficial. In terms of possible directions for improvement:

- The increase of doping in both structures may have been premature. The doping increase originates from the realization of $f_{ul}\tau$ hypothesis of chapter 6 but the impact of band-bending was not understood at the time. Equally good, or even better lasers might result from lowering the doping back to standard levels (at least the power consumption will be lower).

- Improved diagonality was observed to improve performance between OWI223R and OWI219R-M1, so continuing to press in this direction may be beneficial. However, due to the $f_{ul}\tau$ trade-off, at some point increased doping may become necessary.

- Kumar et al. realized a satisfactory 4.7 THz laser in ref. [99] based on a 2-well QCL, although the power dissipation was very high. Optimization of the two-well QCL structure may lead to yet another avenue for high-performance 4.7 THz lasers. The design of ref. [99] was rather vertical ($f_{ul} \sim 0.5$), and one wonders if a greater diagonality might help at such high frequencies.
Figure 8-11: Experimental data for FL182R-M6 (VB0520). Device size is 1.67 mm $\times$ 150 µm. (a) $L-I$ (b) $I-V$ (c) Spectra.
Figure 8-12: Experimental data for FL181R-M7 (VB0518). Device size is 1.87 mm × 150 µm. (a) LI (b) IV (c) Spectra.
8.2 Broadband heterogeneous terahertz QCLs

There has been much recent interest in the deployment of optical combs of all frequencies for high resolution dual-comb spectroscopy. [123] In dual comb spectroscopy, two coherently locked combs of slightly different frequency spacing are mixed together, and the slight differences in frequency results in a corresponding comb at radio or microwave frequencies (∼GHz). [124] In this fashion, very high resolution spectra can be obtained rapidly, limited only by the integration time of any corresponding detectors. Such technology has already been demonstrated in MIR QCLs. [125]

The bandwidth of the spectroscopic measurement is upper-bound by the gain-bandwidth of the laser gain medium. Thus, development of broadband THz QCL gain media is crucial to high performance THz combs. In MIR QCLs, Gmachl et al. have demonstrated that an effective away of achieving broad gain is stacking together QCLs designed for different frequencies. [126] Turcinkova et al. has recently extended this approach to the THz QCL. [127] Motivate by these successes, the broadband gain media of this work are based on heterogeneous stacking as well.

8.2.1 Interlaced heterogeneous terahertz QCL

The design by Turcinkova in [127] consists of three QCLs designs A, B, and C presented in equal proportions layered in the pattern A:C:2B:C:A for a total thickness of about 15 µm. In its original form, this structure suffered from electrical instabilities that prevented the simultaneous lasing of all QCL modules. To overcome these problems, the authors of [127] found it necessary to etch away part of the gain medium in order to restore stability.

Since such instabilities are usually associated with electric field domain formation. Heuristically, a more homogeneous gain medium should be more robust against domain formation, by reason of symmetry. Therefore, in the first attempt to create broadband heterogenous gain medium, the heterogeneous QCL periods were arranged into a superperiod, which was then repeated to produce the appropriate gain medium thickness.
Figure 8-13: Band diagrams of constituent designs of heterogenous gain medium FLR4.

The details are as follows. FL183R, one of the highest power THz QCLs in the literature, was chosen as a base design. This QCL was designed to lase at 4.7 THz (although experimentally it lased at 4.3 THz). Three subsequent designs were made, at 4.2 THz, 3.7 THz, and 3.2 THz. To facilitate the transport of electron between modules of different frequency, all 4 QCL designs were adjusted to have the same design bias (electric field). As the frequency decreased, the LO phonon depopulation gap $E_{21}$ was increased to maintain the same electric field at design bias. The band diagrams for these four design is depicted in figure 8-13.

The designs were tiled in the pattern AD-BC-AD-BC-AD to create a 3:2:2:3 ratio. This ratio of device modules is chosen to achieve a flat gain spectrum (the highest
and lowest frequency designs must be weighted more strongly because of the absence of gain from neighbors). Assuming equal gain from every device, and Lorentzian lineshapes with 1 THz FWHM, the estimated gain spectrum is shown in figure 8-14.

The experimental results are shown in figure 8-15. Although the device achieved a fairly high $T_{\text{max}}$ (142 K), broadband lasing was not achieved. Lasing was apparently completely dominated by design C, with only traces of lasing from A and B, and no lasing at all from D.

### 8.2.2 Charge imbalance in heterogeneous structures

The failure of the interlaced design FLR4 is hypothesized to be caused by charge imbalances caused by mismatches in the current density between different designs. Although the 4 QCL modules were designed to operate at the same field, in practice this might not happen because of current mismatches between the modules. The stack as a whole must always operate with a single current density, and the electric field of each module will adjust itself accordingly.

From Gauss Law, however, a discontinuity in electric field implies the existence
Figure 8-15: Experimental data for heterogenous gain medium FLR4 (VB0621). Device size is 2.12 mm × 150 µm. (a) $LI$ (b) $IV$ (c) Spectra. Lasing is overwhelmingly present at 3.8 THz. Spectra are therefore plotted on logarithmic scale to better highlight weaker lasing features. Lasing at ~4.35 THz, ~4.15 THz, and ~3.8 THz is detected, corresponding to designs FLR4-A, FLR4-B, and FLR4-C, but there is no sign of lasing from FLR4-D.
of a sheet charge at the interface. Thus the QCL modules to either side of a domain boundary between two designs are likely to possess abnormal charge distributions, and hence distorted bandstructures. Hence, the failure of the interlaced design, in which no QCL period is adjacent to a period of the same type, lies in the fact that essentially every QCL module is a boundary module.

8.2.3 Current matching in heterogeneous structures

As stated in section 8.2.2, current density must be constant in a QCL device. However, the IVs of most gain QCL designs are highly nonlinear, with a pronounced negative differential resistance (NDR) at some maximum current density $J_{\text{max}}$. This begs the question of what happens when designs of different maximum $J_{\text{max}}$ are stacked together. In the worst case, a possibility exists that the higher $J_{\text{max}}$ design could force the lower $J_{\text{max}}$ design into NDR as well. In the best case, the constituent designs will not reach their peak biases simultaneously, thus reducing the gain bandwidth. Therefore, $J_{\text{max}}$ should be held constant across all laser designs.

One might think that the ideal solution would be to grow and fabricate each design separately, and adjust each design to match the current densities. This approach is reflected in ref. [128]. However, the problem is that $J_{\text{max}}$ of the different designs must match under lasing conditions. Emphasis has been placed on lasing conditions, because contrary to tradition understanding of THz QCLs, $J_{\text{max}}$ is crucially determined by the presence of lasing. Therefore, the threshold gain of the constituent designs operating in a stack must be known. Unfortunately, the threshold gain of a device tested separately does not reflect the threshold gain of the same device in a heterogenous stack. The threshold gain of a device in stack is increased by the non-unity confinement factor, but is lowered by gain contributed by its neighbors.

8.2.4 Segregated Heterogeneous Gain Media

Given the failure of the interlaced design, a second attempt was made with the more modest goal of simply improving upon the work of ref. [127]. Section 8.2.2 hypoth-
izes that maximally segregated structures are desirable to reduce the interfaces between QCL designs; this change is easily implemented. Section 8.2.3 suggests that the constituent designs should have the same $J_{\text{max}}$ under lasing conditions. This second point is much more difficult to implement.

To implement current matching, the designs of ref. [127] were subjected to an numerical optimization scheme in which sought to maximize the ratio of peak gain to current density under variation of the layer thicknesses and applied electric field. Under this criterion, when all layers were allowed to vary freely, the internal barriers of the QCL module (everything except the injector) were observed to converge to unreasonably thick values. This is due to the physical deficiencies of the model of chapter 3; the model developed in this thesis assumes intramodule quantum coherence between wells separated by arbitrarily thick barriers. Therefore, an alternative approach was adopted in which the barriers were held fixed while the well widths were allow to vary. The band diagrams for the final three designs are shown in figure 8-16.

The predicted $IV$ and gain spectrum of the heterogeneous system are shown in figure 8-17.

Unfortunately, the predictions of figure 8-17 did not happen. The experimental results of OWIE3-M1 are shown in figure 8-18. The resulting laser is quite poor. Although there was no early NDR, the current density is roughly half of that predicted from 8-17 and the lasing is not broadband, being centered mostly around 2.6 THz.

**8.3 Conclusions**

This thesis was unsuccessful in designing broadband heterogeneous gain media. However, some relevant design issues are pointed out.

- The argument is made that active regions should be maximally segregated, to avoid charge build up at the boundaries between different designs.
- Heterogeneous gain medium design is particularly difficult to the the simultane-
Figure 8-16: Band diagrams for constituent designs of OWIE3-M1.
Figure 8-17: Computation $IV$ results for segregated heterogeneous gain medium OWIE3-M1. (a) Gain spectrum under lasing and non-lasing conditions. (b) $IV$ curves for all three designs, and the composite $IV$. 
Figure 8-18: Experimental data for heterogenous gain medium OWIE3-M1 (VB0700). Device size is 1.80 mm × 150 µm. (a) $LI$ (b) $IV$ (c) Spectra. $IV$ shows no sign of early NDR, but has roughly half the current predicted from simulation. Spectra is not broadband, and is centered around 2.6 THz.
ous need to achieve matching current densities between designs and a flat gain spectrum.

An attempted was made to design heterogeneous gain media using the ERE method of chapter 3 but did not succeed. Barring growth errors, the computational tools of this thesis appear to still be deficient for design, at least in this case.
Chapter 9

Summary

9.1 Contributions of this thesis

This thesis represents a concerted attempt on all fronts of theory, computation, and experiment to improve THz QCL maximum lasing temperature. The main contributions of this thesis are as follows.

- A fast numerical solver for THz QCL design was developed, which includes detailed electron-electron scattering and interactions with continuum states.

- A course of systematic waveguide optimization was undertaken, in which the importance of the adhesion layer for waveguide loss is highlighted. A record $T_{\text{max}}$ is achieved using optimized waveguides.

- The $f_{au}\tau$ trade-off is forwarded as an explanation for the failure of diagonality in improving $T_{\text{max}}$.

- Experimental evidence is shown to the effect that the $f_{au}\tau$ trade-off can be defeated using increased doping. However, accounting for band-bending effects becomes important in highly doped designs.

- Computational and experimental evidence is presented to suggest that leakage to continuum and shunt leakages via high energy subbands is not an important degradation mechanism in current THz QCLs.
9.2 Directions for future work

I am often asked if a room temperature Terahertz Quantum Cascade laser is possible at all. There is no answer to this question, but I believe there is still hope in the strategy laid out in this thesis. Much work remains to be done. Most importantly, the defeat of the $f_{ul}\tau$ trade-off through increased doping has found some experimental support, but the hypothesis on degradation through band-bending at high doping is as yet unproven.

My recommendation is to pursue the doping strategy in two-well THz QCLs, where band-bending misalignment is unimportant. If a sustained rise in $T_{\text{max}}$ versus doping can be demonstrated in two-well designs, such a result would be significant, even if no $T_{\text{max}}$ record is broken; it would suggest promise in the strategy of carefully accounting for Poisson effects in standard resonant phonon designs.

Accounting for the such Poisson effects may take two forms: an experimental approach might be to test several THz QCLs designed under different assumptions about the electron distribution (for example, different splits among the subbands), and seeing which assumption works best. However, the range of possible electron distributions is extremely wide, and such a pursuit might be costly.

The theoretical approach would be to continue improving the modeling of THz QCLs. I hope the effective rate equations (ERE) model of this thesis will be useful to future work. The utility of the ERE model is that is captures the vast majority of the physics available to ensemble Monte Carlo simulations at a fraction of the computational expense. Unfortunately, while in many cases it yields results in excellent agreement with experiment, but it also fails badly in many others.

Most of the ERE model’s shortcomings are fundamental, such as the artificial division between coherent transport and scattering, the reliance on a tight-binding basis, and the inability to calculate dephasing. Minor improvements in physics such as implementing state-blocking and nonparabolicity are unlikely to produce substantially different results. Moving to a more complete method such as nonequilibrium Green’s functions (NEGF) may be necessary.
Analogous to the ERE model being a faster method of capturing the physics of ensemble Monte Carlo, perhaps a second simulator could be made with the aim of being a faster version of NEGF. The ERE achieves its computational efficiency by making use of the assumption that the subbands in a THz QCL are essentially thermal, which enables the use of analytic forms for the distribution functions. Applying this strategy of analytical approximation to NEGF might be profitable. Furthermore, a real-space basis, similar to ref. [46], may be considered if detailed electron-electron scattering is investigated. The self-energies for electron-electron scattering are much more simply expressed in real-space compared to eigenstates of the energy.
Bibliography


[95] A. Tredicucci. communicated to Q. Hu during a visit to MIT.


[111] M. A. Belkin, K. Vijayraghavan, A. Vizbaras, A. Jiang, F. Demmerle, G. Boehm, R. Meyer, M. C. Amann, A. Matyas, R. Chashmahcharagh, P. Lugli,


Appendix A

$k \cdot p$ parameters

Table A.1 lists the $k \cdot p$ used in this work for $\text{Al}_x\text{Ga}_{1-x}\text{As}$. Parameters are generally taken from the work of Vurgaftman et al. [112]. Parameter interpolation was performed by linearly interpolating between the effective masses, and then working backwards to extract the $k \cdot p$ parameters.

The most notable exception to the parameter set of Vurgaftman is that in most cases a conduction band offset of $1.247 \cdot 0.72x$ eV is used instead of Vurgaftman’s recommendations. This is historical value used for design in the THz-MMW group at MIT. Vurgaftman’s recommendations were, however, followed for the tall barrier designs of chapter 7, where the $1.247 \cdot 0.72x$ eV formula clearly fails. Explicitly, the conduction band offset given by Vurgaftman is

\[
\Delta E_c = E_{g,\text{AlGaAs}} - E_{g,\text{GaAs}} - \Delta E_v
\]
\[
= 1.519 (1 - x) + 3.099x - x (1 - x) (-0.127 + 1.310x) - 1.519 - 0.53x \quad (A.1)
\]
\[
= 1.310x^3 - 1.437x^2 + 1.177x
\]

Despite the technological maturity of the AlGaAs material system, there is still a wide range of conduction band offsets used in the literature. For example, researchers at the University of Waterloo use $0.65(1.36 + 0.22x).$ [129] Figure A-1 plots various offset formulae versus aluminum fraction $x$ for comparison.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value in AlGaAs</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_g$</td>
<td>Bandgap</td>
<td>$1.519 (1 - x) + 3.099 x - x (1 - x) (-0.127 + 1.310 x)$</td>
<td>eV</td>
</tr>
<tr>
<td>$E_c$</td>
<td>Conduction bandedge</td>
<td>$1.247 \cdot 0.72 x, \ x \sim 0.15$</td>
<td>eV</td>
</tr>
<tr>
<td>$E_v$</td>
<td>Valence bandedge</td>
<td>$E_c - E_g$</td>
<td>eV</td>
</tr>
<tr>
<td>$\Delta_{so}$</td>
<td>Split-off bandedge</td>
<td>$0.341 (1 - x) + 0.28 x$</td>
<td>eV</td>
</tr>
<tr>
<td>$E_{so}$</td>
<td>Split-off bandedge</td>
<td>$E_v - \Delta_{so}$</td>
<td>eV</td>
</tr>
<tr>
<td>$m_e$</td>
<td>Electron effective mass</td>
<td>$0.067(1 - x) + 0.15 x$</td>
<td>$m_0$</td>
</tr>
<tr>
<td>$m_{lh}$</td>
<td>Light-hole effective mass</td>
<td>$0.1852 (1 - x) + 0.0901 x$</td>
<td>$m_0$</td>
</tr>
<tr>
<td>$m_{hh}$</td>
<td>Heavy-hole effective mass</td>
<td>$0.4717 (1 - x) + 0.3497 x$</td>
<td>$m_0$</td>
</tr>
<tr>
<td>$F_K$</td>
<td>Kane parameter</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td>$E_P = 2mP^2$</td>
<td>Kane energy</td>
<td>$\frac{E_x (E_x + \Delta_{so})}{E_g + 2 \Delta_{so} / 3} \left( \frac{m_0}{m_e} - 1 - F_K \right)$</td>
<td>eV</td>
</tr>
<tr>
<td>$\gamma_1^L$</td>
<td>Luttinger’s 1st parameter</td>
<td>$\frac{1}{2} \left( \frac{m_0}{m_{lh}} + \frac{m_0}{m_{hh}} \right)$</td>
<td></td>
</tr>
<tr>
<td>$\gamma_2^L$</td>
<td>Luttinger’s 2nd parameter</td>
<td>$\frac{1}{4} \left( \frac{m_0}{m_{lh}} - \frac{m_0}{m_{hh}} \right)$</td>
<td></td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>Modified Lutt.’s 1st parameter</td>
<td>$\gamma_1^L - \frac{E_P}{3E_g}$</td>
<td></td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>Modified Lutt.’s 2nd parameter</td>
<td>$\gamma_2^L - \frac{E_P}{6E_g}$</td>
<td></td>
</tr>
</tbody>
</table>

Table A.1: $k \cdot p$ parameters used for Al$_x$Ga$_{1-x}$As
Figure A-1: Γ-point conduction band offset for Al\(_x\)Ga\(_{1-x}\)As/GaAs as a function of Al fraction \(x\), from refs. [73], [112] and [129] respectively.
### Appendix B

## Sample growth sheet

Growth sheet is for OWI210H-M3 from chapter 6.

<table>
<thead>
<tr>
<th>250°C</th>
<th>GaAs</th>
<th>LTG (250°C) cap layer</th>
<th>35Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>GaAs</td>
<td>$5 \times 10^{19}$ cm$^{-3}$</td>
<td>100 Å</td>
</tr>
<tr>
<td>Contact</td>
<td>GaAs</td>
<td>$5 \times 10^{18}$ cm$^{-3}$</td>
<td>500 Å</td>
</tr>
<tr>
<td>Injector</td>
<td>Al$<em>{0.15}$Ga$</em>{0.85}$As</td>
<td>undoped</td>
<td>48.0 Å</td>
</tr>
<tr>
<td>Injector</td>
<td>GaAs</td>
<td>undoped</td>
<td>59.3 Å</td>
</tr>
<tr>
<td>Injector</td>
<td>GaAs</td>
<td>$6.2 \times 10^{16}$ cm$^{-3}$</td>
<td>48.0 Å</td>
</tr>
<tr>
<td>Injector</td>
<td>GaAs</td>
<td>undoped</td>
<td>59.3 Å</td>
</tr>
<tr>
<td>Injector</td>
<td>Al$<em>{0.15}$Ga$</em>{0.85}$As</td>
<td>undoped</td>
<td>56.5 Å</td>
</tr>
<tr>
<td>Repeat 210 times</td>
<td>GaAs</td>
<td>undoped</td>
<td>84.8 Å</td>
</tr>
<tr>
<td>Repeat 210 times</td>
<td>Al$<em>{0.15}$Ga$</em>{0.85}$As</td>
<td>undoped</td>
<td>36.7 Å</td>
</tr>
<tr>
<td>Repeat 210 times</td>
<td>GaAs</td>
<td>undoped</td>
<td>87.6 Å</td>
</tr>
<tr>
<td>Repeat 210 times</td>
<td>Al$<em>{0.15}$Ga$</em>{0.85}$As</td>
<td>undoped</td>
<td>48.0 Å</td>
</tr>
<tr>
<td>Repeat 210 times</td>
<td>GaAs</td>
<td>undoped</td>
<td>59.3 Å</td>
</tr>
<tr>
<td>Repeat 210 times</td>
<td>GaAs</td>
<td>$6.2 \times 10^{16}$ cm$^{-3}$</td>
<td>48.0 Å</td>
</tr>
<tr>
<td>Repeat 210 times</td>
<td>GaAs</td>
<td>undoped</td>
<td>59.3 Å</td>
</tr>
<tr>
<td>Repeat 210 times</td>
<td>Al$<em>{0.15}$Ga$</em>{0.85}$As</td>
<td>undoped</td>
<td>56.5 Å</td>
</tr>
<tr>
<td>Contact</td>
<td>GaAs</td>
<td>$5 \times 10^{18}$ cm$^{-3}$</td>
<td>0.1 µm</td>
</tr>
<tr>
<td>Etch-stop</td>
<td>Al$<em>{0.15}$Ga$</em>{0.85}$As</td>
<td>undoped</td>
<td>0.4 µm</td>
</tr>
</tbody>
</table>

$n^+$-GaAs substrate
Appendix C

Additional processing information

This is a compilation of other cleanroom processing information that may be helpful to other users in replicating this work.

Selective etch of Au on GaAs

During the course of this thesis, on at least one occasion there was a need to recover samples deposited with contaminated Ti/Au. Unfortunately, most substances which etch Au also etch GaAs. Woodruff et al., however, found that KI:I$_2$:H$_2$O gold etchants acidified with HCl selectively etched gold off Ge nanowires. [130] This method was subsequently extended to GaAs nanowires by Liu. [131] Pursuant to the work of ref. [131], a 9:1 solution of Transene gold etchant TFA (produced by Transene Corp.) to concentrated (37%) HCl was found to satisfactorily removed Au selective to GaAs, with at least $\sim$ 15:1 selectivity (the GaAs etch depth was hard to assess due to its shallowness). The adhesion layer (Ti/Ta) can then be removed with HF.

Photoresist masks for dry etching

Occasionally, photoresists masks were used for etching dielectric films, but often the photoresist film would become burnt during the etch and become unremovable. This is because of poor thermal contact between the carrier plate/wafer and the sample.
This can be alleviated by performing dry-etches for only a minute at a time, with 1-min breaks in between. Following etch, the photoresist will nevertheless have a hard crust on the outside that does not wash off with acetone. Following an acetone rinse, the remaining crust can be removed by ashing.

**Photoresist adhesion for wet etching**

During the course of this thesis, photoresist adhesion failures were frequently encountered, particularly during deep wet etching (the photoresists used were Shipley 1813 and AZ5124E). A number of pre-coating treatments were considered to solve these adhesion problems, including

1. A 3-solvent clean (acetone followed by methanol followed by isopropanol).

2. A 5 min descum in O\(_2\) plasma (asher).

3. Treatment with Hexamethyldisilazane (HMDS), as is frequently done with Si wafers.


A simple 3-solvent clean was found to yield the best adhesion results. Ashing and HMDS yielded no improvement at all, and the oxide strip actually caused severe adhesion loss. Based on these results, consistently good adhesion was achieved by performing a solvent clean immediately prior to spin coating, and ensuring wet chemical cleans and etches were separated from spin-coating steps by at least 12 hours (ie. typically, wet etches and spin-coating are separated by an overnight wait period.)

The fact that the oxide strip actively ruined photoresist adhesion was extremely surprising, as this procedure was specifically recommended in one publication for improving adhesion. \[^{132}\] The lesson to take away is that fabrication recipes are not necessarily portable between cleanrooms. Literature recipes must be carefully tested before production use.
Defect reduction in Au-Au bonding

As mentioned in section 4.1.2, Cu-Cu bonding anecdotally seems to have the fewest bonding defects. In some circumstances though, Au-Au bonding may be preferable, such as if a vacuum environment is not available to prevent Cu oxidation. Some success in improving Au-Au bonding quality was achieved through a pre-bonding treatment consisting of a 10 min ash followed by a 30 s dip in HCl.