Title
Design and Modeling of High Temperature Terahertz Quantum Cascade Lasers

Permalink
https://escholarship.org/uc/item/9bx7p5kf

Author
Burnett, Benjamin Adams

Publication Date
2016

Peer reviewed|Thesis/dissertation
Design and Modeling of High Temperature Terahertz Quantum Cascade Lasers

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Electrical Engineering

by

Benjamin Adams Burnett

2016
Abstract of the Dissertation

Design and Modeling of High Temperature Terahertz Quantum Cascade Lasers

by

Benjamin Adams Burnett

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2016

Professor Benjamin S. Williams, Chair

The portion of the electromagnetic spectrum between roughly 300 GHz and 10 THz is nicknamed the “THz Gap” because of the enormous difficulty encountered by researchers to devise practical sources covering it. Still, the quantum cascade laser (QCL) has emerged over recent years as the most promising approach to a practical source in the 1-5 THz range. First developed in the higher-frequency mid-IR, where they are now widely available, QCLs were later extended to the THz where a host of greater design challenges awaited. Lasing in QCLs is based on intersubband optical transitions in semiconductor quantum wells, the energy of which can be chosen by design (“bandstructure engineering”). However, simply building a THz optical transition is insufficient; a good design must also produce significant population inversion by the applied cascading electron current, and this requires deep understanding of the transport physics. So far, no THz QCL has operated above the temperature of 200 K, even though the reasons prohibiting high temperature operation are well known.

The goal of this Thesis is to put novel ideas for high-temperature operation of THz QCL active regions through rigorous theoretical testing. The central enabling development is a density-matrix-based model of transport and optical properties tailored for use in QCLs, which is general enough that widely varying design concepts can be tested using the same core principles. Importantly, by simulating QCLs more generally, fewer a priori assumptions are required on part of the researcher, allowing for the true physics to emerge on its own.
It will be shown that this gives rise to new and useful insights that will help to guide the experimental efforts towards realization of these devices.

One specific application is a quantum dot cascade laser (QDCL), a highly ambitious approach in which the electrons cascade through a series of quantum dots rather than wells. Benefits are expected due to the suppression of nonradiative scattering, brought about by the discrete spectrum of electronic states. However, this in turn leads to a highly different physics of transport and effects that are not well understood, even in the case of perfect materials. This work will show that while the benefits are clear, naïve scaling of existing QCL designs to the quantum dot limit will not work. An alternative strategy is given based on a revised understanding of the nature of transport, and is put to a test of practicality in which the effects of quantum dot size inhomogeneity are estimated.

Another application is to the already existing method of THz difference frequency generation in mid-IR QCLs, which occurs via a difference-frequency susceptibility $\chi^{(2)}$ in the active region itself. For this purpose, the model is extended to enable a coherent and non-perturbative calculation of optical nonlinearities. First, the generality of the method is displayed through the emergence of exotic nonlinear effects, including electromagnetically-induced transparency, in mock quantum-well systems. Then, the modeling concepts are applied to the real devices, where two new and important mechanisms contributing to $\chi^{(2)}$ are identified. Most importantly, it is predicted that the QCL acts as an extremely fast photodetector of itself, giving rise to a current response to the mid-IR beatnote that provides a better path forward to the generation of frequencies below $\sim 2$ THz.

Finally, the fundamentals of density matrix transport theory for QCLs are revisited to develop a model for conventional THz QCL designs eliminating the usual phenomenological treatment of scattering. The new theory is fully developed from first principles, and in particular sheds light on the effects of scattering-induced electron localization. The versatility of the model is demonstrated by successful simulation of varying active region designs.
The dissertation of Benjamin Adams Burnett is approved.

Yaroslav Tserkovnyak
Oscar M. Stafsudd
Diana L. Huffaker
Benjamin S. Williams, Committee Chair

University of California, Los Angeles
2016
To my family
# Table of Contents

1 Introduction ......................................................... 1
   1.1 Quantum Cascade Lasers .................................. 4
      1.1.1 Active Regions .................................. 5
      1.1.2 Waveguides ...................................... 6
   1.2 The Temperature Problem in THz QCLs ...................... 8
      1.2.1 Origin .......................................... 8
      1.2.2 Quantum dot cascade lasers ...................... 10
      1.2.3 Terahertz difference frequency generation in mid-infrared QCLs ... 13
   1.3 Overview .................................................. 16

2 QCL Bandstructure and Transport Modeling ...................... 17
   2.1 Bandstructure .............................................. 17
      2.1.1 Effective mass model ................................ 17
      2.1.2 Subbands in quantum wells .......................... 20
      2.1.3 Sublevels in quantum dots ......................... 22
      2.1.4 Nonparabolicity: Three-band Kane bandstructure model ........... 24
   2.2 Optical selection rules and oscillator strengths .......... 27
      2.2.1 Quantum wells .................................... 28
      2.2.2 Quantum dots .................................... 29
      2.2.3 Nonparabolicity .................................. 30
   2.3 Overview of existing QCL transport models ............... 30
   2.4 Density matrix transport modeling ......................... 32
      2.4.1 The density matrix ................................ 32
4 Realistic Design Strategy For a Quantum Dot Cascade Laser  . . . . . . . 72
  4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 72
  4.2 Design concept . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
    4.2.1 Key polaron physics . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
    4.2.2 Design strategy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75
  4.3 Extension of the model to three phonon modes . . . . . . . . . . . . . . . . 78
  4.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
  4.5 Effects of inhomogeneity . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
  4.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84

5 Terahertz Difference Frequency Susceptibility in Midinfrared Quantum
Cascade Lasers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 86
  5.1 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
    5.1.1 Resonant susceptibility . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
    5.1.2 Previous theoretical work . . . . . . . . . . . . . . . . . . . . . . . . . 89
  5.2 Extension of density matrix concepts for nonlinear optics . . . . . . . . . . 90
    5.2.1 Newly identified effects for THz difference frequency generation . . . 90
    5.2.2 Steady-state equations . . . . . . . . . . . . . . . . . . . . . . . . . . . 92
    5.2.3 Velocity extraction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93
    5.2.4 Example: Two-level system with permanent dipole . . . . . . . . . . 94
    5.2.5 Example: Three-level system . . . . . . . . . . . . . . . . . . . . . . . 96
    5.2.6 Example: Electromagnetically-induced transparency . . . . . . . . . . 97
  5.3 Scattering calculations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101
    5.3.1 LO-phonons . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103
    5.3.2 Interface roughness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3.6</td>
<td>Agreement with Fermi’s Golden Rule</td>
<td>147</td>
</tr>
<tr>
<td>6.4</td>
<td>Example: superlattice tunneling</td>
<td>148</td>
</tr>
<tr>
<td>6.5</td>
<td>Application to various QCLs</td>
<td>151</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Diagonal hybrid design at ( \sim 3.4 ) THz</td>
<td>152</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Vertical design at ( \sim 3.0 ) THz</td>
<td>155</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Vertical design at ( \sim 2.1 ) THz</td>
<td>157</td>
</tr>
<tr>
<td>6.5.4</td>
<td>Scattering-assisted design at ( \sim 1.8 ) THz</td>
<td>158</td>
</tr>
<tr>
<td>6.6</td>
<td>Conclusions</td>
<td>160</td>
</tr>
<tr>
<td>7</td>
<td>Conclusions</td>
<td>162</td>
</tr>
<tr>
<td>A</td>
<td>Steady-state density matrix solution and velocity extraction</td>
<td>167</td>
</tr>
<tr>
<td>A.1</td>
<td>Elementwise</td>
<td>167</td>
</tr>
<tr>
<td>A.2</td>
<td>Density matrix vectorization</td>
<td>170</td>
</tr>
<tr>
<td>B</td>
<td>Evaluation of tunnel couplings under nonparabolicity</td>
<td>176</td>
</tr>
<tr>
<td>C</td>
<td>Future experiment: nanopillar intraband photoconductor</td>
<td>178</td>
</tr>
<tr>
<td>C.1</td>
<td>Design concept</td>
<td>178</td>
</tr>
<tr>
<td>C.2</td>
<td>Electromagnetic simulations</td>
<td>180</td>
</tr>
<tr>
<td>C.3</td>
<td>Fabrication details</td>
<td>182</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>184</td>
</tr>
</tbody>
</table>
1.1 Operating temperatures of quantum cascade lasers in the mid-infrared and THz. Figure from [41]. ......................................................... 3
1.2 Conduction band structures for example mid-infrared and THz QCL active regions. Mid-IR design from [44], THz design adapted from [34]. Material systems (well/barrier) are In$_{0.53}$Ga$_{0.47}$As/In$_{0.52}$Al$_{0.48}$As for the mid-infrared and GaAs/Al$_{0.15}$Ga$_{0.85}$As for the THz. ......................................................... 5
1.3 Waveguide configurations used in THz QCLs. (a) Semi-insulating surface plasmon (SISP) waveguide, (b) Metal-metal waveguide. Figure from [45]. ......................................................... 6
1.4 Depiction of LO-phonon emission in mid-IR and THz transitions. The green represents a thermal electron distribution in the upper subband, and the red arrows show the change in an electron’s state for example LO-phonon emission processes. ......................................................... 8
1.5 Renditions of (left) a quantum cascade laser, and (right) a quantum dot cascade laser. Drawings are not to scale. Note that the use of self-assembled QDs in a QDCL would not yield inherent vertical alignment as drawn. ......................................................... 10
1.6 (left) SEM images of top-down etched quantum cascade active material. Taken from [73]. (right) SEM Images of MOCVD-grown nanopillars. left: In content in a GaAs/InGaAs axial heterostructure as a function of height, taken using line-scan energy dispersive X-ray spectroscopy (from Ref. [34]). top right: InAs pillars grown in select locations, image courtesy of Huffaker Group. bottom right: InAs pillars planarized with benzocyclobutene (BCB) polymer and etched to reveal tips for device processing, also courtesy of Huffaker Group. ......................................................... 11
1.7 Progression of the THz peak power record by nonlinear difference-frequency generation in mid-IR QCLs, since invention of the technique in 2007. All records shown are at room temperature, except for 2007. Figure taken from [92]. ......................................................... 14
2.1 Paraboloidal subband energy structure in quantum wells. The energy minima depend on the heterostructure sequence. .................................................. 22

2.2 Comparison of effective mass and three-band Kane bandstructure models for (left) a well for typical mid-IR subband separation and (right) a well for typical THz subband separation. The conduction band components of the wavefunctions are plotted at the eigenenergy positions. ................................. 26

2.3 Basis choices around a resonant tunneling barrier. ............................................. 35

2.4 An example of periodicity in a QCL active region, with interactions occurring both inside and between modules. .......................................................... 39

3.1 Schematic and time dynamics of relaxation by electron-phonon Rabi oscillations and eventual phonon decay. ............................................................... 47

3.2 The particular phonon mode interacting with a cylindrical QD lowest-lying transition: (left) upward and (right) radial displacements. Both are in separate arbitrary units, and the radial displacement is shown at a phase $\pi/2$ relative to that of the upward. ......................................................... 50

3.3 Orthonormalization within the mode space of two modes $N$ and $R$: the orthogonal modes are $N$ and $\alpha$. .............................................................. 53

3.4 LO-phonon decay and generation as transition rates between number (Fock) states of a mode. ................................................................. 56

3.5 Bandstructure in the growth direction for the model system, computed from a two-well tight-binding standpoint. .................................................. 59

3.6 Relevant tunneling and electron-phonon interactions in the electron-phonon tensor product basis. ............................................................... 61

3.7 Computed gain profiles under vanishing optical intensity at the injection anticrossing bias for different pure dephasing times, at 100 K and 300 K. Colored numbers on the left are inverted population fractions. ......................... 63

3.8 Gain profile separated into phonon occupations for $T = 300$ K and $T_2^* = \infty$. 64
3.9 Zero-phonon gain at 300 K and $T_2^* = \infty$ as the tunnel coupling is turned on. Curves are offset by 200 cm$^{-1}$ for clarity. Thin lines denote $E_{rad} \pm \Omega_{tun}$ and $E_{rad} \pm \Omega_{tun} \pm \sqrt{\Omega_{tun}^2 + \Omega_{pol,N}^2}$.

3.10 Gain at 300 K, $T_2^* = 5$ ps as a destructive scattering mechanism is introduced. Legend denotes values of $\tau_{sp}$. 

3.11 Gain saturation as optical intensity is increased, for $T = 300$ K, $T_2^* = 5$ ps.

3.12 Effects of a parasitic tunneling channel at $T = 300$ K. Left: Transport characteristic for $T_2^* = 5$ ps and 1 ps. Right: Gain at various bias points for $T_2^* = 5$ ps. (labeled in mV/module and offset by 200 cm$^{-1}$ for clarity).

4.1 (a) Energy eigenstates obtained with (solid) and without (dashed) the polaron coupling in an InAs QD with varying sublevel separations $E_2 - E_1$. $\Omega_{pol}$ is given a typical value of 3 meV. Downward arrows depict start/end states of electron relaxation. (b,c) Relaxation dynamics for different sublevel energy separations assuming the LO-phonon decay time of 2.5 ps and for pure dephasing times of $T_2^* = \infty, 3$ ps. Colors correspond to downward arrows in (c).

4.2 Electronic bandstructures and level schematics at parasitic/design biases for (a) a conventional QCL design and (b) a modified design for a QDCL. The chosen material system is InAs/InAsP and the layer thickness in nm starting with the tunnel barrier are 5.5/14/3/30 for (a) and 5.5/18/3/35 for (b), so that the only difference is a widening of the wells. The level schematics illustrate the design strategy which is to engineer resonant LO-phonon depopulation of the upper state $\Psi_U$, rather than the lower state $\Psi_L$, to the injector state $\Psi_I$.

4.3 Schematic of electron-phonon interactions in the tensor product Hilbert space when three phonon modes are considered. Orange arrows labeled by Roman numerals represent groups of couplings.
4.4 Simulated transport characteristics and bias-dependent small-signal optical gain at 300 K for the conventional and modified designs, showing regions of electrical stability/instability. The designed and predicted gain in the modified design is circled by the dashed oval.

4.5 Simulated effects of saturation on the conventional and modified designs at their respective design biases. Saturated gain profiles in (a) and (b) are cross-saturation by intensity at $\omega_{\text{peak}}$ (marked by dashed line), and (c) is the peak of these profiles in relation to the small-signal gain at $\omega_{\text{peak}}$.

4.6 Simulated gain spectra of 50 different variations of each design, where all length dimensions vary by a Gaussian distribution of 5% standard deviation. The solid black lines are the gain in the intended designs, and the solid red lines are the average of all variations.

5.1 Bandstructures in an example mid-IR QCL designed to emit THz radiation by difference frequency generation. The highlighted states participate the most in lasing action.

5.2 Depiction of the resonant difference frequency susceptibility present in Active Region B (Fig. 5.1), occurring across the inverted optical transitions.

5.3 Illustration of effects that will be shown to be important for THz difference frequency generation in mid-IR QCLs. Mechanisms outlined in red have yet to see thorough theoretical description.

5.4 The scheme used to evaluate Equation 5.7.

5.5 (Top) schematic of a two-level system with permanent dipole. (Bottom) Difference frequency susceptibility calculated at (left) a fixed signal of 10 meV, (right) a fixed pump of 100 meV.
5.6 (Top) Schematic, assumed parameters, and calculated parameters of the three-level system created by coupling an antisymmetric well pair by a $\delta$-potential. 
(Bottom) Calculated difference frequency susceptibility with added pure dephasing $T_2^* = 300$ fs. 

5.7 Emergence of the EIT effect as more frequencies are included in the steady-state solution.

5.8 Effect of permanent dipoles on EIT, comparing the steady-state solution with and without diagonals of $Z$ included, for different intensity levels.

5.9 Example LO-phonon scattering times in Active Region A.

5.10 Example interface roughness scattering times in Active Region A.

5.11 Example alloy disorder scattering times in Active Region A.

5.12 Example ionized impurity scattering times in Active Region A.

5.13 Assumed electron distributions for selected subbands in Active Region B.

5.14 (a) Transport characteristics for Active Regions A and B. (b) Gain spectra and saturation for Active Regions A and B at bias points such that $J = 10$ kA/cm$^2$. 

xv
5.15 (a,b) Simulated difference frequency current susceptibility $Y^{(2)}(\omega_1 = \omega_3 - \omega_2)$ with increasing pump intensity for (a) Active Region A (long wavelength) and (b) Active Region B (short wavelength). The difference frequency $\omega_1$ is swept with pump $\omega_2$ fixed at 130 meV and the other pump $\omega_3$ swept in conjunction. The black, blue, and green lines correspond to 0.1 kW/mm$^2$, 1 kW/mm$^2$, and 10 kW/mm$^2$, respectively, in each pump. Values at optical rectification are marked by the dots. (c,d) The increase in current as a function of a stimulating frequency as found from a single-frequency simulation for (c) the long wavelength and (d) the short wavelength active regions. (Input intensities are four times the intensity of the single fields in a,b.) Scaling by $2nc$ reproduces the value of $Y^{(2)}$ at optical rectification (denoted by dots), although only for small intensity. Equivalent points in the two simulations along the horizontal axis (single frequency input at 130 meV) are marked by the dashed lines, and resonant processes are denoted on each plot.

5.16 Simulated increase in current density for both active regions in response to a pump at energy 130 meV. The intensity axis is extended to 40 kW/mm$^2$, the peak intensity when beating two pumps of 10 kW/mm$^2$ each. Linear expansion of each around zero intensity is given by the dashed lines.

5.17 Comparison of $Y^{(2)}(\omega_1 = \omega_3 - \omega_2)$ in Active Region B for the full calculation vs. a sum-over-states for three different scenarios: (a) swept-signal generation with one pump fixed at 130 meV, (b) fixed-signal generation of 4 THz, (c) fixed-signal generation of 1 THz. The magnitudes are shown in the main plots and phases are displayed in the insets. The calculations were done for vanishing intensity. Units for $|Y^{(2)}|$ are Mm/Vs.
5.18 Fitting of the full calculation to a simpler model involving superposition of detection and SOS components. The black, red, and blue lines give results from the full calculation, SOS expression with permanent dipoles, and SOS expression without permanent dipoles. Calculations are performed on Active Region B, for the same fixed-pump scenario as in Figs. 5.15a,b and 5.17a. Real and imaginary parts are given for vanishing intensity in (a) and (b), and for 10 kW/mm² in each pump in (c) and (d). The fitted parameter $\beta = 0.110 ((\text{kA/cm}^2)/(\text{kW/mm}^2))$ for vanishing intensity, although at 10 kW/mm² in each pump this coefficient is reduced to 0.065 $((\text{kA/cm}^2)/(\text{kW/mm}^2))$. The approximate response time $\tau$ is 240 fs.

5.19 Effect of the injection barrier width (in legends) on $Y^{(2)}$ for vanishing intensity. Green lines correspond to values chosen in Ref. [44].

6.1 LO-phonon emission rates extracted from the coarse-gridded superoperator, compared with Fermi’s Golden Rule.

6.2 Bandstructures in the superlattice system of two levels per period, with increasing barrier thickness but constant well width. Anticrossing biases are chosen. The highlighted states form one module.

6.3 (left axis) Current in the two-level superlattice with increasing barrier width, for the full density matrix model and a semiclassical model with the coherences eliminated. (right axis) Electron center with respect to the middle of the barrier.

6.4 Bandstructure and simulated results for a diagonal hybrid design in GaAs/Al$_{0.15}$G$_{0.85}$As, aimed at $\sim$ 3.4 THz. The layer thicknesses in nm starting at the tunnel barrier are 5.1/10.3/1.7/10.7/3.7/8.8/3.7/17.2 and a sheet doping of $4.38 \times 10^{10}$ cm$^{-2}$ is assumed at the center of the wide well. The lattice temperature was set to 100 K.
6.5 Bandstructure and simulated results for a vertical design in GaAs/Al$_{0.15}$G$_{0.85}$As, aimed at $\sim$ 3.0 THz. The layer thicknesses in nm starting at the tunnel barrier are 4.9/7.9/2.5/6.6/4.1/15.6/3.3/9.0 and a sheet doping of $3 \times 10^{10}$ cm$^{-2}$ is assumed at the center of the wide well. The lattice temperature was set to 100 K.

6.6 Bandstructure and simulated results for a vertical design in GaAs/Al$_{0.15}$G$_{0.85}$As, aimed at $\sim$ 2.1 THz. The layer thicknesses in nm starting at the tunnel barrier are 5.6/8.2/3.1/7.0/4.2/16.0/3.4/9.6 and a sheet doping of $3.4 \times 10^{10}$ cm$^{-2}$ is assumed at the center of the wide well. The lattice temperature was set to 77 K.

6.7 Bandstructure and simulated results for a diagonal $\sim$ 4.0 THz/scattering-assisted $\sim$ 1.8 THz design in GaAs/Al$_{0.15}$G$_{0.85}$As. The layer thicknesses in nm starting at the tunnel barrier are 4.2/8.5/2.3/9.6/3.4/7.3/4.0/15.3 and a sheet doping of $3 \times 10^{10}$ cm$^{-2}$ is assumed at the center of the wide well. The lattice temperature was set to 100 K.

B.1 Accuracy of the method for calculation of tunnel couplings under strong non-parabolicity.

C.1 NIP device concept (not to scale).

C.2 Simulated results for the surface plasmon grating coupler concept.

C.3 NIP fabrication process overview.
Acknowledgments

First, I would like to thank my advisor, Professor Benjamin Williams, for giving me the opportunity to work on such exciting projects, and for serving as a top-notch mentor along the way. My graduate work journeyed through places I think neither of us ever imagined when I began, but as things unfolded, Professor Williams was supportive of the ideas I had, and had the patience to guide me in the right direction as I developed my own understanding of the field. I went to graduate school really hoping to become a physicist in the context of engineering, and that is exactly what I got.

I also want to thank the other members of my committee: Professors Diana Huffaker, Oscar Stafsudd, and Yaroslav Tserkovnyak. A special thanks goes to Professor Huffaker as our long-time collaborator on the experimental front. Students in Professor Huffaker’s group who worked on the nanopillar growth efforts are Joshua Shapiro, Andrew Lin, Adam Scofield, and Dingkun Ren, from whom I learned much about the world of MOCVD that you just can’t really get from the publications. Thanks also goes out to Alan Farrell, who was instrumental in teaching me fabrication techniques and helping me develop the process for a nanopillar intraband photoconductor.

Andrew Pan has been a solid collaborator, fellow theorist, and friend. What started out as casual discussions about our work and feedback on manuscripts eventually became a full-blown research effort into a truly generalized density matrix transport model for QCLs. Over the years, Andrew has helped me to grasp a lot of the deeper aspects of transport physics, and for the most part this occurred in the cafeteria at the Reagan hospital.

Of course, thanks goes out to my own former and present labmates: in particular Amir Tavallaee, Philip Hon, Zhijun Liu, Luyao Xu, and Chris Curwen. We’ve always had a friendly, engaging, and smart group of people in here, starting with Amir who was a founding member and continuing beyond me. So much knowledge gets passed down between the generations, and never have I seen anyone hesitate to lend a helping hand to pay the favors forward. For that reason, I should again thank Amir, Philip, and Zhijun for helping me when I was
starting out.

Some former mentors deserve credit too. Thanks to Professor Joshua Semeter at BU, with whom I got my start and initial passion for research. Professor Semeter was kind enough to take me on as an undergrad researcher, and went above and beyond to ensure that it was a rich experience. I also want to thank the following teachers from high school who contributed to my development as a scientist way back then: Mr. Griffin, Mrs. Walla, Mr. Pflieger, and Mrs. Ohanessian.

I want to thank all of my friends both inside and outside of UCLA who were there as supporters over a very long period of life. Without these people, I would have lost my mind, and I am blessed to have kept strong friendships with people from everywhere I’ve gone, all the way back to early childhood. I can’t begin to name them because there are too many (I’m a popular guy), but an enormous thank you to all just for being my friend and joining me in the things I love.

Finally, the biggest thank you goes to my family and to Zrinka. Mom and Dad instilled in me a passion for learning and asking questions that took me to where I am today and will stick with me for the rest of my life. They always valued education, so I guess it’s not surprising I did a PhD, but what they really did for me during this time was their offering of support through all the challenges. The same of course goes for Davey, the best big brother I could ask for and an inspiration since the day I was born. My last thank you is to Zrinka, who has been right there with me for just about this entire thing, be it the easy times or the hard times. That alone says worlds about her, but she’s a lot of fun too.
Facilities This work has benefited from the use of two cleanrooms at UCLA: the Nanoelectronics Research Facility and the Integrated Systems Nanofabrication Cleanroom, of which the support staff were excellent. Some of the computing work was performed using the compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The staff at the CHTC were extremely helpful and knowledgeable, and delivered an excellent OSG Summer School in Madison where I learned to use their resources.

Funding We thank the National Science Foundation for funding support throughout this work.
VITA

2006-2010  B.S. in Electrical Engineering with Physics Minor and Nanotechnology Concentration, Boston University

2008  Electronics Engineering Intern, HR Textron, Santa Clarita, California

2008-2010  Undergraduate research assistant under Professor Joshua Semeter, Boston University Electrical and Computer Engineering

2010-present  Graduate Research Assistant under Professor Benjamin Williams, UCLA Electrical Engineering

2012  M.S. in Electrical Engineering, UCLA

2013-2015  Teaching Assistant, UCLA Electrical Engineering

PUBLICATIONS

B.A. Burnett and B.S. Williams, “Design strategy for terahertz quantum dot cascade lasers,” submitted manuscript (2016).


B.A. Burnett and B.S. Williams, “Density matrix modeling of terahertz difference frequency generation in dual active region quantum cascade lasers”, The 13th International Conference on Intersubband Transitions in Quantum Wells (Vienna, Austria), September 7th, 2015.

CHAPTER 1

Introduction

Sources of electromagnetic radiation can typically be classified as either electronic or photonic. In both cases, the radiation is generated by an electric current, but it is the nature of this current that sets the difference: electronic sources radiate by conduction currents, while photonic sources radiate by displacement currents. The lower-frequency electronic sources are therefore limited by the speed of electronic circuits, while the higher-frequency photonic sources are limited by the available material properties. It is unfortunate that the traditional capabilities of both sides do not meet in the middle, and so the range in between (∼300 GHz-10 THz) has become known as the “Terahertz Gap”.

Despite the difficulty in producing practical coherent THz sources, sensitive detectors have existed for quite some time [1, 2], making it possible and worthwhile to pursue THz applications while source technologies mature [3, 4]. A major motivator for THz technology is in astronomy, as blackbody radiation from extremely cold interstellar dust is peaked in the THz, so much so that it is estimated that 98% of photons and half the luminosity in the universe lie within this range [5, 6]. Recent major astronomical research experiments involving THz sensing include the satellite-based Herschel Space Observatory [7], SOFIA (Stratospheric Observatory for Infrared Astronomy), a space telescope mounted inside a 747 airliner [8], and STO (Stratospheric Terahertz Observatory), a balloon mission. All three projects boast multiple detectors in various parts of the THz range, and several of these utilize sources for heterodyne detection, including quantum cascade lasers for the GREAT [9] and upGREAT [10] receivers on SOFIA. Other THz applications outside of astronomy include security imaging [11], detection of explosives and illicit drugs [12, 13], biomedical imaging by hydration sensing [14], and even non-destructive evaluation of precious artworks.
There have been many successful approaches to terahertz generation throughout the years, although all have their respective drawbacks. As the scope of this thesis is primarily the part of the THz range faster than 1 THz, we will focus on these approaches. Electronic-type sources include high electron mobility transistors (HEMTs) and double heterojunction bipolar transistors (DHBTs) [16, 17], but these experience high-frequency roll-off proportional to $f^{-4}$, due to RC and transit time limitations. So far only one non-extrapolated demonstration of amplifier gain at $f_{\text{max}} > 1$ THz has been made [18]. There are also devices based on the use of negative differential resistance in resonant tunneling diodes (RTDs) to cancel the loss of LC resonators [19, 20], but these experience similar limitations. Other sources perform upconversion from the microwave, for instance using Schottky diode multiplier chains [21], but in this case the conversion efficiency is low, and worsens with the output frequency such that above 2 THz only microwatts are possible.

Bridging electronic- and optical-type sources is photoconductive terahertz generation, which is based on optically-induced modulation of the carrier density in semiconductors [22, 23]. These devices have now achieved up to 17 $\mu$W in quasi-cw operation (50% duty cycle) at 1 THz [24]. However, the electronic character of these devices causes similar high-frequency roll-off, making them most ideal in the sub-1 THz range, such as for pulsed sources in THz time-domain spectroscopy.

Truly photonic-type sources include “far-infrared” (i.e. THz) gas lasers, which can produce watt-level powers at select wavelengths but are large and bulky, especially considering the extremely powerful pump lasers [25]. Far-infrared lasing has also been demonstrated in p-type Ge assisted by either very large magnetic fields [26] or compressive strain [27], making the full system also quite bulky. It has been possible to downconvert from the infrared and optical ranges using a nonlinear crystal placed inside a high power laser cavity, but in this case the system size is dominated by the high power laser, and the low conversion efficiencies limit the state-of-the-art systems to $\sim 1$ mW of THz radiation from several hundred watts of intracavity pump intensity [28, 29]. Fitting into neither the electronic- nor optical-type categories are free-electron lasers (FELs), which are building-sized facilities in which an electron
beam passes through a magnetic undulator. FELs have produced much important scientific research by delivering extremely high powers at inaccessible wavelengths [30, 31], but given the facility size and cost cannot be used in practical applications.

Finally, the source at the focal point of this thesis is the quantum cascade laser (QCL), the present leading candidate for a practical source in the 1-5 THz range. QCLs are a photonic-type source where stimulated emission occurs between artificially-engineered “subbands” that are spaced at a chosen photon energy, entirely within the semiconductor’s conduction band. Pumping is electrical, making for a nicely compact package. QCLs were actually first demonstrated in the mid-infrared in 1994 [32], and it was not until 2002 that they were extended to the longer THz wavelengths [33]. In the time since, the record-setting demonstrations include 1.01 W peak power in pulsed operation (10 K) [34] and 230 mW cw power (10 K) [35, 36], although the highest-temperature operation is still only 200 K [37]. Therefore, while powers are easily high enough for real-world applications, the operating temperature is the primary inhibitor to commercial adoption.
Progress on the temperature front has slowed in recent years; almost a decade ago in 2008, the highest reported temperature was not far lower at 178 K [38, 39]. This suggests that existing approaches are close to optimized, and that novel ideas are required if the THz QCL is to get the boost to room temperature, or even that of thermoelectric coolers. It is interesting to note that this was not the case for mid-infrared QCLs, which quickly progressed beyond room temperature and have since moved on to widespread commercialization [40], a position envied by the THz counterpart as visualized in Figure 1.1. In this thesis, I will explore the physics of popular strategies for high-temperature operation in THz QCL active regions, aiming for a strong physical understanding leading to new and robust design strategies.

### 1.1 Quantum Cascade Lasers

QCLs owe their existence to modern epitaxial growth technologies, typically Molecular Beam Epitaxy (MBE), but also including Metal-Organic Chemical Vapor Deposition (MOCVD) [42, 43]. These systems can deposit ultrathin semiconductor layers of varying composition at near atomic-layer precision, over total thicknesses up to 10 µm. If the thicknesses are sufficiently small (at least to the order of tens of nm in III-V materials), quantum confinement results in the splitting of the bands into “subbands” whose energies and other properties depend intricately on the chosen material sequence. Aided by detailed knowledge of the material properties, the bandstructure can be artificially engineered, which has led to an enormous range of technologies over the past few decades.

QCLs and certain other quantum-well devices such as Quantum Well Infrared Photodetectors (QWIPs) are made to be unipolar by n-doping of the material, thus eliminating the participation of the valence bands. This greatly reduces the complexity of the problem, as the relevant subband structure becomes only that of the conduction band, making design strategies easier to devise. If two layers of varying bandgap are deposited, the resultant effect on the conduction band edge energy is a modulation, i.e. formation of quantum “wells” and “barriers”, since electrons sit at lower energy in the lower bandgap material. Freedom
of choice in the well and barrier thicknesses is a vast design space, which has led to QCLs and QWIPs operating over two orders of magnitude in the electromagnetic spectrum, with a key distinction between the mid-IR and THz regions, naturally made by the material’s Reststrahlen band (\(\sim 8-9\) THz in most III-V materials).

The general strategy for both mid-IR and THz QCL active regions is to engineer a sequence of a few wells and barriers which is repeated a large number (often 100s) of times. An electrical bias applied across the whole active region causes a tilt in the conduction band edge and also drives current through the device. Lasing action is possible if the electrons build up at the upper state of an optical transition within each module so as to provide stimulated emission. The entire active region is encased in a metallic waveguide serving as a cavity, whose actual design depends on the specific device goals.

### 1.1.1 Active Regions

The ability to design the operating frequency in QCLs stems from basic quantum mechanical principles, that by confining electrons to tighter spaces the allowed energies become more widely separated. Figure 1.2 shows the bandstructure for example mid-IR and THz QCL
active regions, where the functions shown are the probability density in the growth direction for the subbands, placed at the subband energy minimum. In both cases, the goal is to have a population inversion leading to stimulated emission: this requires a larger population in the “upper” subbands at the top of the optical transition than in the “lower”. This is achieved by engineering fast injection to the upper subbands and fast extraction from the lower subbands, but also requires the upper state lifetime to be sufficiently long to cause a population buildup in the upper subbands. Transport occurs through a multitude of tunneling and scattering processes, which are discussed and calculated in great detail throughout this thesis. Intuitively, the vast difference in energy scale between the two spectral regimes causes somewhat different physics to dominate, ultimately leading to the stark contrast in performances between the two.

1.1.2 Waveguides

Metallic waveguides surrounding QCLs both provide electrical contact to the active region and also serve as a cavity, with the end mirrors usually formed simply by the index contrast between semiconductor and air and the impedance mismatch of the mode. Subwavelength confinement of the mode causes the far-field beam pattern to be highly divergent, the im-
provement of which is an active area of research.

Since a major source of loss in QCLs is the free-carrier (Drude) loss in conductors, it has been necessary to devise schemes to minimize the amount of lasing mode field inside or tangential to metal or doped layers. Low-doped layers are especially damaging to THz, rather than mid-IR, QCLs because the loss above the plasma frequency scales as $\lambda^2$. Figure 1.3 shows the two most commonly used waveguide configurations for THz QCLs: semi-insulating surface plasmon waveguides (SISP) and metal-metal (MM) waveguides. SISP waveguides use a highly-doped but thin layer as the bottom contact, which sits on top of a semi-insulating substrate. The lasing mode is a surface plasmon mode hosted on the doped layer, extending both upwards into the active region and downwards into the substrate. SISP waveguides allow for the highest output powers because of lower loss, by virtue of the mode being held further away from the metal. In addition, the larger mode area radiates a less divergent beam pattern. MM waveguides, on the other hand, use metal for both top and bottom contacts, exhibiting a higher confinement factor of the lasing mode to the active region. This is advantageous for higher temperature operation, since the modal overlap factor with the active region is larger as compared to SISP waveguides. MM waveguides can also be made much narrower than the wavelength, which makes for more efficient heat removal [46, 47] and also enables certain methods of tuning by altering the properties of the lasing mode that extends well outside the cavity [48]. SISP waveguides cannot be made nearly as narrow, as there is a minimum width within which the mode leaks almost entirely into the substrate [49].

Since MM waveguides are highly desirable for their temperature performance, there have been a number of approaches to improving the highly divergent beam patterns. Methods of direct collimation have included the use of a lens [50] or horn antenna [51]. There have also been approaches relying on increasing the size of the radiating aperture using emission distributed across the whole laser cavity, such as by a 3rd order distributed-feedback (DFB) grating [52], or using vertical emission (vertical direction in Fig. 1.3) from a photonic crystal cavity [53], microring second-order DFB [54], or from arrays of ridges which share a cavity [55]. Recently, a THz QC-VECSEL (vertical external cavity surface emitting laser)
was demonstrated in which arrays of ridges with high loss to free space act as an amplifying metasurface inside of a larger cavity [56]. This approach lases in a highly Gaussian fundamental mode, with only a few degrees of FWHM divergence.

### 1.2 The Temperature Problem in THz QCLs

#### 1.2.1 Origin

It is now fairly well agreed upon that the dominant force against high temperature operation is thermally-activated longitudinal-optical (LO-) phonon scattering [45, 57, 58, 59, 60]. LO-phonons are strong inelastic scatterers with a narrow energy window, at $E_{LO} = 36$ meV (8.7 THz) in GaAs and in the same vicinity for other III-V materials. It is this energy selectivity that gives LO-phonons such a strong dependence on electron temperature which the other most relevant scattering mechanisms do not exhibit.

To understand the thermal dependence of the LO-phonon scattering process, we consider two subbands at either mid-IR or THz energy separation shown in Figure 1.4, which would represent the upper and lower lasing subbands in a QCL. Supposing there exists a population inversion of the upper subband to the lower one, we will approximate that electrons are distributed thermally within the upper subband, due to fast intrasubband scattering. For
the mid-IR transition, LO-phonon emission is energetically allowed from any state in the upper subband, although a large momentum transfer is required. As will be shown in detail in Section ???, this somewhat reduces the strength of LO-phonon scattering across mid-IR transitions. On the other hand, in a THz transition LO-phonon emission is not energetically allowed from the upper subband except for electrons at an energy of at least $E_{LO}$ above the lower subband energy minimum. However, for electrons above this threshold the LO-phonon scattering is extremely fast (often sub-picosecond), since a relatively low momentum transfer is needed. If the electron temperature is sufficiently cold, the vast majority of electrons sit below the energy threshold, suppressing the effect, but with rising temperature the thermal spread ($k_B T$) reaches beyond it, amounting to a thermally activated process. If we make the approximations of Boltzmann-distributed subbands and equal scattering rate for electrons above the minimum energy, then the subband-averaged transition rate $W_{i \rightarrow f}$ for subbands spaced at $E_{if} < E_{LO}$ is:

$$W_{i \rightarrow f} = W_{if}^{(\text{hot})} e^{-(E_{LO} - E_{if})/k_B T}. \quad (1.1)$$

By this equation we see that the proximity of the thermal energy $k_B T = 26$ meV at room temperature to the LO-phonon energy $E_{LO}$ and to the THz energies is the reason that such thermal activation occurs.

Other factors have been put forward that contribute to performance degradation with temperature, but none have the extreme temperature dependence governing the LO-phonon scattering. One factor is the thermal backfilling of the laser’s lower state, which is that as the device warms up a larger number of electrons climb energetically back up into this state, where they reduce the gain [45]. However, with the advent of resonant-phonon QCL designs, the energy separation between the lower state and the state below it is around $E_{LO}$, which is greater than the thermal spread $k_B T$ at room temperature, making this effect fairly small. Additionally, this is not expected to affect the upper state population strongly. Another factor is the leakage current through parasitic states sitting energetically high in the wells or continuum states above the barrier [59], which can also be approximated as having
an activation energy which is the energy from the upper state to these places. Given the
typical barrier heights greater than 100 meV, however, this effect is expected also to remain
relatively small. Another factor is the increased waveguide loss due to shortened Drude
scattering times at higher temperature, but this effect is also not so strongly temperature
dependent. Finally, increased broadening due to increased impurity scattering at higher
temperature has also been suggested as a factor in gain reduction with temperature [57].

1.2.2 Quantum dot cascade lasers

Even though LO-phonon scattering occurs in only the narrow energy window around $E_{LO}$,
as explained in the previous section LO-phonons are able to induce transitions between
subbands of energy separation not equal to $E_{LO}$ because of the continuum of electronic
states above each subband minimum, and the momentum carried by the LO-phonons. This
loosening of constraints opens up a wide set of LO-phonon-induced transitions which obey
both energy and momentum conservation. Since the first is enabled by the existence of
continuous electronic states, one can readily imagine that with discrete states this would not
be the case. Such a laser has been envisioned in the QCL community, where the discrete
density of electronic states present in quantum dots (QDs) could create misalignment between
electron transitions and the LO-phonon energy, hopefully mitigating the detrimental LO-
phonon scattering [61, 62, 63]. This effect is commonly referred to in the literature as
“phonon bottleneck” [64].
Figure 1.6: (left) SEM images of top-down etched quantum cascade active material. Taken from [73]. (right) SEM Images of MOCVD-grown nanopillars. left: In content in a GaAs/InGaAs axial heterostructure as a function of height, taken using line-scan energy dispersive X-ray spectroscopy (from Ref. [34]). top right: InAs pillars grown in select locations, image courtesy of Huffaker Group, bottom right: InAs pillars planarized with benzocyclobutene (BCB) polymer and etched to reveal tips for device processing, also courtesy of Huffaker Group.

Such a device, known as a quantum dot cascade laser (QDCL), remains far from demonstrated. Figure 1.5 shows a rendition of a QDCL in comparison to an ordinary QCL. While in the vertical direction the active material in a QDCL would follow a similar sequence, the material is separated into nanopillars which confine the electrons in the in-plane directions to a quantum mechanical degree. The effect on the optical properties of the active region is not problematic, since the desired spacings (order of 100 nm) are far subwavelength to THz fields, giving rise to an “effective” homogeneous medium. Given varying ways to planarize nanopillars to make top electrical contact, waveguide fabrication in a QDCL could proceed using similar techniques as in ordinary QCLs.

By far the most challenging aspect of the realization of a QDCL is the production of the active material. Attempts have been made using self-assembled quantum dots (SAQDs), top-down etching of QCL active material, and bottom-up growth. Images of the latter two are in Figure 1.6. The SAQD approach has seen intense theoretical study [65, 66], but on the experimental side the furthest advancement is electroluminescence from QDs embedded in quantum cascade material [67, 68, 69, 70], including at room temperature [71], and THz
luminescence activated by a free-electron laser [72]. In such a design, the QDs would ideally have lower energy than the well such that the QD ground (s) state has an energy just above the bottom of the well. The intended lasing transition occurs from the excited (p) state to this state, from where it would in principle be extracted by additional bandstructure design. However, the lack of vertical alignment between dots in different layers makes this difficult, and there is also a large amount of inhomogeneous broadening due to the inherent stochastic nature of the self-assembly process. Top-down etching of planar QC material has produced broadband electroluminescence over the THz and mid-IR in pillars of 160 nm diameter [73], and also THz lasing at 77 K from arrays of 5 µm pillars, which is significant because the pillar spacing is subwavelength to the THz field [74]. Difficulties in top-down etching include the challenge of producing good sidewalls, and also the high density of surface states which exists because the pillar walls do not follow a crystal plane, causing both surface-charge-induced depletion of the pillars and scattering.

Bottom-up growth of a QDCL would involve the same technologies as ordinary QCLs, namely MBE or MOCVD. Under certain growth conditions, nanopillars can naturally form in the growth reactor, seeded either by a catalyst such as a gold droplet [75] or by small holes of substrate in select locations exposed beneath an oxide layer. Site control, sometimes referred to as “selective-area epitaxy,” is possible in both catalyzed and non-catalyzed growth using electron-beam [76, 77, 78, 79] or nano-imprint [80] lithographies. While bottom-up growth tends to yield close to perfectly-hexagonal sidewalls defined by crystal planes, control in the vertical direction is still at its early stages [81, 79, 82, 83], in contrast to the near atomic precision possible with quantum wells. Therefore, it can be said that top-down etching boasts better control in the vertical direction, while bottom-up growth is better in the in-plane directions.

Although a true QDCL has not been realized, a number of experimental demonstrations have motivated the concept as the costly growth technology develops. By breaking subbands into discrete Landau levels, a magnetic field creates a discretization effect similar to lateral confinement, and in one experiment this increased the operating temperature of a THz QCL from 160 K to 225 K, for a field strength of 19 T where the Landau level separation is
approximately $E_{LO}$ [84]. Furthermore, it had been previously shown that fluctuations in
the output power of a THz QCL occur in response to an increasing magnetic field as the
Landau levels move in and out of resonance with $E_{LO}$ [85]. In actual QDs, the phonon
bottleneck effect has been measured directly using pump-probe measurements of the excited
state lifetimes of InGaAs SAQDs with varying size: at 77K, the lifetime of a 14.5 meV
transition was three orders of magnitude longer than that of a 29.5 meV transition (near
$E_{LO}$) [86].

In addition to the experimental challenges, the theory behind QDCLs is not entirely
complete. The replacement of an electronic state continuum with discrete states causes
fundamentally different physics to emerge, perhaps most importantly that the LO-phonon
scattering is replaced by the formation of long-lasting electron-phonon quasiparticles named
“polarons” [87]. This will be the subject matter of Chapter 3 and a review of the existing
theoretical work is given there.

1.2.3 Terahertz difference frequency generation in mid-infrared QCLs

Although coherent THz radiation has never been generated in a QCL at room temperature
directly, it has in fact been accomplished using indirect nonlinear processes. The key, which
was noticed even before the first QCL demonstration, is that quantum wells can possess
optical nonlinearities much higher than occur in natural materials, owing to the large dipole
matrix elements [88, 89, 90, 91]. For example, difference frequency susceptibilities ($\chi^{(2)}$)
are estimated at tens of nm/V, in contrast to only 1-300 pm/V found in typical nonlinear
crystals [92].

Making use of this so-called “giant nonlinear response,” researchers have invented the
clever idea to harness the power and maturity of mid-IR QCLs to generate THz radiation at
room temperature, circumventing the temperature issues that plague THz QCL designs [93].
Two mid-IR active regions are sandwiched in the same cavity and biased in series, and the
cavity is designed to allow for simultaneous operation at two different mid-IR wavelengths.
The very quantum well heterostructures that produce amplification of the mid-IR waves
then also act to mix the two frequencies, generating a polarization wave at the difference frequency between the two, which is placed by design in the THz. This polarization wave radiates inside the cavity, yielding some amount of THz power in addition to the mid-IR beams. Since the active region is able to produce the mid-IR pumps at room temperature, the resulting THz radiation at room temperature is achieved.

Ever since this concept was first demonstrated at room temperature in 2008 [94], operating powers have advanced at an almost exponential rate (see Fig. 1.7), owing to a better understanding of the nonlinear susceptibility and the phase matching/loss considerations. A very meaningful expression for the generated THz power inside the waveguide \( P_1 \) is:

\[
P_1 = \frac{l_{coh}^2 P_2 P_3}{8 \varepsilon_0 c^3 n_1 n_2 n_3 S_{eff}} \omega_1^2 |\chi^{(2)}|^2,
\]

where \( P_{2,3} \) are the mid-infrared pump powers, \( l_{coh} \) is the coherence length of the nonlinear interaction, \( n_{1,2,3} \) are refractive indeces, \( S_{eff} \) is the effective area of interaction, and \( \chi^{(2)} \) is the difference frequency susceptibility for generation of \( \omega_1 = \omega_3 - \omega_2 \). While \( \chi^{(2)} \) in these devices is far greater than for conventional nonlinear generation, the coherence length is in
fact much worse. The coherence length involves both loss of the generated THz wave and also phase matching, which are both poor inside of a mid-IR QCL cavity; phase-matching is poor because placement of the Reststrahlen band between the THz and mid-IR causes their respective refractive indices to be quite different, and THz losses are also quite high in cavities designed to provide strong mid-IR pumps. There is somewhat of a trade-off involved in selection of the doping density, as a higher doping is desired for higher mid-IR pump power and $\chi^{(2)}$ nonlinearity, but at the same time increases the THz loss. A scheme has been demonstrated to solve the phase-matching problem, by use of Cerenkov radiation where the generated THz beam is automatically phase-matched at an angle pointing downwards into the substrate, owing to a faster phase velocity of the polarization wave in the active medium than the generated wave in the substrate [95]. Although this approach has now been widely used and leads to the most successful devices, the issue of THz loss still stands. On the $\chi^{(2)}$ front, the same strategy as was used in the first demonstration has not changed, which is to optimize a particular resonant susceptibility [93] in the active region. This is generally associated with the mid-IR lasing transition, giving the added advantage that the pumps experience gain rather than loss, in contrast to typical nonlinear generation.

Although the record powers by nonlinear generation still lag those of cryogenically-cooled THz QCLs by a few orders of magnitude, progression in the area has been steady; the highest peak THz power in pulsed mode is now 1.9 mW [96], and the cw record is 14 $\mu$W [97]. Already, however, difference frequency generation dominates the QCL field in tunability range, with the broadest tuning spectrum being all the way from 1.2-5.9 THz [98]. This is accomplished using an external Littrow cavity which is configured so that one pump frequency is set while the other can be tuned within the gain bandwidth [44]. The extremely broad frequency range is possible because only small relative tuning in the mid-IR produces enormous relative tuning across the THz. Intriguingly, the low-frequency cutoff is caused by a switch to single-mode operation due to gain saturation, rather than a loss of nonlinear response; even in the face of increased Drude losses, the $\omega_1^2$ dependence in Eq. 1.2, and movement away from the expected $\chi^{(2)}$ resonance. Explanations for this surprising behavior are put forward in Chapter 5.
1.3 Overview

This thesis is focused on the theoretical modeling of novel concepts aimed at high temperature operation in THz QCLs, enabled by a density matrix formalism that was developed to be widely applicable to QCLs. Chapter 2 gives an overview of the relevant transport physics, and an outline of the density matrix approach that is employed in various forms throughout the thesis. In Chapter 3, the modeling ideas are applied to a THz quantum dot cascade laser (QDCL), and the expected important physics is reviewed there. Serious ramifications for the operation of THz QDCLs are identified and their quantitative strength predicted. Chapter 4 builds upon the insight gained in Chapter 3, and suggests new and counterintuitive design strategies for QDCLs which are quantitatively tested. Chapter 5 extends the density matrix transport modeling concepts to optical nonlinearity, where fresh insight is gained into the mechanisms of THz difference frequency susceptibility in mid-IR QCLs. The newly identified mechanisms suggest strategies for their optimization, which is expected to be useful particularly for target frequencies slower than 2 THz. Finally, Chapter 6 outlines a way to generalize the inclusion of incoherent scattering processes in the density matrix model, which eliminates the need for a phenomenological basis choice and is in fact necessary for the modeling of certain THz QCL active regions.
CHAPTER 2

QCL Bandstructure and Transport Modeling

2.1 Bandstructure

2.1.1 Effective mass model

The electronic energy eigenstates in a periodic potential such as a semiconductor are the Bloch states, whose wavefunctions are:

\[ \Psi = \frac{1}{\sqrt{S_{nk}}} u_{nk} e^{i\vec{k} \cdot \vec{r}}. \]  

(2.1)

\( \vec{k} \) is the electron wavevector, also known as the crystal momentum, and \( u_{nk} \) is the Bloch function, a periodic function following the lattice which can in general vary with both \( \vec{k} \) and the band index \( n \). The factor \( S_{nk} \) is a normalization. The Bloch wave is expressible as a Floquet series, since it is periodic with the added complexity that there is a constant phase shift between neighboring unit cells. Often in band theory, the Bloch function in one band at different wavevectors \( \vec{k} \) is expanded in terms of the Bloch functions of all bands at some high symmetry point, which for the purposes here will be the \( \Gamma \)-point (\( \vec{k} = 0 \)). Under such an expansion, the wavefunction above becomes:

\[ \Psi = \sum_m u_{m0} c_{mn} e^{i\vec{k} \cdot \vec{r}}, \]  

(2.2)

where \( u_{m0} \) are the \( \Gamma \)-point Bloch functions in all bands \( m \) and \( c_{mn} \) are coefficients which now incorporate the normalization. In this representation, the electron wavefunctions away from the \( \Gamma \)-point have a certain “character” in all the other bands, as we can assign an envelope...
function $F_m = c_{mn}e^{ik \cdot r}$ to each particular Bloch function $u_{m0}$. It is for the specific case of a bulk material that the envelope functions are plane waves, but for heterostructures the envelope function must be taken generally, explicitly:

$$\Psi = \sum_m F_m u_m,$$

(2.3)

where $F_m$ are the envelope functions associated with each $\Gamma$-point Bloch function and $u_m$ is a shorthand notation now introduced for $u_{m0}$. Assuming this form of the energy eigenstates, it will be shown that a Schrödinger-like eigenvalue equation can be obtained for the envelope functions and energies.

Application of the full Schrödinger equation to the wavefunction above yields the following:

$$\sum_m \left[ \frac{p^2}{2m_0} + V - E \right] F_m u_m = 0,$$

(2.4)

where $V$ is the full crystal potential including the heterostructure. Evaluating the $p^2$ operator leads to:

$$\sum_m \left[ \frac{F_m p^2 u_m}{2m_0} + \frac{u_m p^2 F_m}{2m_0} + \frac{(pF_m) \cdot (pu_m)}{m_0} + (V - E) F_m u_m \right] = 0.$$

(2.5)

If the basis functions $u_m$ are piecewise the Bloch functions in the respective materials, then the Schrödinger equation gives that $p^2 u_m + V u_m = E_m u_m$, where $E_m$ is the $\Gamma$-point energy of band $m$. A full analysis resulting in proper ordering of the operators in the end actually uses the $u_m$ basis functions as the same in different materials [99, 100], but to more simply illustrate the point we will proceed in this way and identify the correct ordering at the end.

Using the relation in the previous paragraph and integrating Eq. 2.5 multiplied with another basis function $u_n$ over a unit cell leads to:
\[
\left( \frac{p^2}{2m_0} + E_m - E \right) F_n + \frac{1}{m_0} \sum_{m \neq n} (pF_m) \cdot p_{nm} = 0,
\]

(2.6)

under the approximation that the envelope functions are slowly varying compared to the lattice constant. \(p_{nm} = \langle u_n | p | u_m \rangle\) are the momentum matrix elements for the Bloch functions.

If we are interested in solutions near the Γ-point of one particular band, then these solutions will primarily have the \(u_m\) of that band. This band will be called "c" since in this thesis it will always be the conduction band, and all the other "remote" bands will be labeled with indeces \(r\). The strategy in calculating the bandstructure around this energy is the same as is used in traditional \(k \cdot p\) theory, which is to fully consider the conduction band while perturbatively adding the effect of all the others. When energies are far enough away from the Γ-point that some bands can no longer be added perturbatively, a nice solution is provided by Kane’s theory in which the Hamiltonian is directly solved using the important bands, while the rest still act perturbatively [101]. A three-band model using this method is outlined in Section 2.1.4.

If the \(F_n\) in Eq. 2.6 is of one of the remote bands, then we make the approximations (1) that the kinetic energy in this band is negligible in comparison to the distance \(E_r - E\) between the eigenstate and the remote band minimum, and (2) that influence on this envelope function from other bands (last term) comes primarily from the conduction band since the other remote band \(F_r\) functions are small. This leads to:

\[
F_r = \frac{1}{m_0} \frac{(pF_c) \cdot p_{rc}}{E_m - E}.
\]

(2.7)

If \(F_n\) in Eq. 2.6 is the conduction band, we have:

\[
\left( \frac{p^2}{2m_0} + E_c - E \right) F_c + \frac{1}{m_0} \sum_r (pF_r) \cdot p_{cr} = 0.
\]

(2.8)

Substituting Eq. 2.7 into 2.8 and being somewhat loose with the ordering of operators (justified later) leads to an eigenvalue equation for \(F_c\):

19
\[
\frac{p^2}{2m_0} F_c + \frac{1}{m_0^2} \sum_r p_r p_c \frac{E_r - E}{E} p F_c + E_c F_c = E F_c. \tag{2.9}
\]

This is now in a Schrödinger-like form with a position and energy-dependent effective mass tensor \(m^*(E, \vec{r})\). However, if we assume the tensor \(p_r p_c\) to be isotropic, it reduces to a constant \(\|p_c\|^2\), and if we assume the remote bands to be far away in energy such that \(E_r - E \approx E_r - E_c\), then the equation further reduces to one having a scalar mass with only position dependence:

\[
p \frac{1}{2m^*(\vec{r})} p F_c + E_c F_c = E F_c \tag{2.10}
\]

\[
\frac{1}{m^*(\vec{r})} = \frac{1}{m_0} + \frac{2}{m_0^2} \sum_r \frac{|p_c|^2}{E_r - E_c}. \tag{2.11}
\]

Eq. 2.10 is known as the effective-mass Schrödinger equation. The ordering of operators is important since doing it this way guarantees the continuity of both \(F_c\) and \(\frac{1}{m^*} \nabla \cdot F_c\), which are necessary for current continuity. The effective mass \(m^*\) is usually gained from experiment rather than theory and is often very different from \(m_0\) (\(m^* = 0.067m_0\) in GaAs, \(0.023m_0\) in InAs).

The effective-mass Schrödinger equation is quite powerful because it means that we are able bury the effect of the periodic lattice into the effective mass \(m^*\) with the only added complexity being its position dependence. In a sense, the problem has started in Schrödinger’s equation, moved through the complexity of band theory, and in the end returned to a simple Schrödinger form. This means that semiconductor heterostructures can be designed using intuitive concepts taken directly from basic quantum mechanics.

### 2.1.2 Subbands in quantum wells

In quantum wells, the material, and thus \(E_c\), are varying in the one growth direction \(z\) but are held constant in the other. Therefore, since the potential is separable into \(z\) and in-plane
components \((E_c = V_z(z) + V_\parallel(x,y))\), we perform the usual separation of variables used in quantum mechanics:

\begin{align*}
F_c &= \psi_z(z)\psi_\parallel(x,y) \\
E &= E_z + E_\parallel.
\end{align*}

(2.12) (2.13)

Applying the effective-mass Schrödinger equation and giving only \(z\) dependence to \(m^*\) leads to two separate equations for the different components:

\begin{align*}
-\hbar^2 \frac{d}{dz} \frac{1}{m^*(z)} \frac{d}{dz} \psi_z + V_z \psi_z &= E_z \psi_z \\
-\hbar^2 \frac{1}{2m^*(z)} \nabla_\parallel^2 \psi_\parallel + V_\parallel \psi_\parallel &= E_\parallel \psi_\parallel,
\end{align*}

(2.14) (2.15)

where \(\nabla_\parallel^2\) is the Laplacian operator in only the two in-plane directions. We see that the spatial-dependence of the effective mass actually somewhat couples the \(z\) and in-plane directions, working against the separation of variables assumption, but it is common to ignore this entering into Eq. 2.15. The solutions to \(E_z\) are then some discrete set depending on the actual quantum well sequence, and for \(V_\parallel = 0\) the \(\psi_\parallel\) are plane waves with the continuous spectrum \(E_\parallel = \hbar^2 k_\parallel^2 / 2m^*\). Therefore, solutions can be organized into “subbands”, which have paraboloidal structure in \(k_\parallel\)-space and energy minima given by the solutions \(E_z\). A spatially-dependent effective mass is used to obtain these energies and the growth-direction wavefunctions \(\psi_z\), but for the in-plane dispersion it is typically assumed that the effective mass is simply the mass in the well, justified by the fact that the overall wavefunctions tend to be much more localized there. Figure 2.1 shows an example of subband energy structure.

Although the in-plane problem has a trivial analytical solution, the one-dimensional growth direction problem must in general be solved numerically. One way to do this is to discretize the effective mass Schrödinger equation (2.15) onto points \(z_n\) spaced by \(\Delta_z\) [102]:

21
Figure 2.1: Paraboloidal subband energy structure in quantum wells. The energy minima depend on the heterostructure sequence.

\[ -s_n \psi_{n-1} + d_n \psi_n - s_{n+1} \psi_{n+1} = E \psi_n \]  
(2.16)

\[ s_n = \frac{\hbar^2}{\Delta^2 (m^*_n - 1 + m^*_n)} \]  
(2.17)

\[ d_n = \frac{\hbar^2}{\Delta^2} \left( \frac{1}{m^*_{n-1} + m^*_n} + \frac{1}{m^*_n + m^*_{n+1}} \right) + V_n \]  
(2.18)

For this method it is important to choose the grid points such that interfaces are midway between. This equation can then be solved for an arbitrary potential by finding the matrix eigenvalues. Usually, the boundary condition \( \psi_1 = \psi_N = 0 \) is applied, with \( N \) being the number of spatial grid points.

2.1.3 Sublevels in quantum dots

Eq. 2.10 could be solved directly to get the bandstructure in quantum dots, but this would be very computationally intensive and so it is usually advantageous to take advantage of some sort of symmetry. Since in this thesis we treat quantum dots in a pillar geometry, if we assume that the pillar boundary presents an infinite potential then a separation of variables between the vertical and in-plane directions is still valid. Therefore, we can continue to use Eqs. 2.14 and 2.15 and will make the same assumption that \( m^*(z) = m^*_{\text{well}} \) for the in-plane
problem to allow for a separated solution.

The solution in the growth direction can then follow in the same way as for quantum wells, whereas now the in-plane problem is no longer so trivial as the plane-wave solutions. Certain special cases are exactly solvable; for example if the pillar cross-section is a box or a circle. The unnormalized solutions in a circle are Bessel functions:

$$\psi_{2D}(\rho, \phi) = J_{|m|}(k\rho)e^{im\phi}, \quad (2.19)$$

which are angular momentum eigenstates having $L_z = m\hbar$, $m$ being the angular momentum quantum number which can run from 0 to $\infty$. $k$ is the radial wavevector, and it is restricted by a boundary condition that the wavefunction must be zero at the pillar boundary, giving the allowed values of $k$ and the energies as:

$$k = \sqrt{\frac{2m^*E_\parallel}{\hbar}} = \frac{a_{m,n_r}}{R}. \quad (2.20)$$

$R$ is the pillar radius, $n_r$ is the radial quantum number, and $a_{m,n_r}$ is the $(n_r + 1)^{th}$ zero of the $m^{th}$ order Bessel function. Thus, the solutions can be indexed by the two quantum numbers $m$ and $n_r$ which both can run from 0 to $\infty$ without selection rules on their allowed combinations. The energy of a state is dependent on both numbers, and when sorted in energy a plot of them is close to linear, reflecting the flat density of states for bulk 2D materials. Although realistic semiconductor nanopillars have actually a hexagonal, rather than circular, cross-section, the cylindrical symmetry can often be exploited for calculations as will be clear throughout this thesis, and in fact the eigenergies are quite close to those obtained by fully solving Schrödinger’s equation in a hexagonal potential [103].

If a general solution to the effective mass Schrödinger equation in two dimensions is desired, it is possible to do so using the same discretization scheme as for the one-dimensional case in the previous section. Neglecting the spatial dependence of the effective mass and introducing a 2D spatial grid indexed by $n$ and $m$, the discretized form is:
\[-\frac{\hbar^2}{2m^*\Delta^2}(\psi_{n,m+1} + \psi_{n,m-1} + \psi_{n+1,m} - 4\psi_{n,m}) + V_{nm}\psi_{nm} = E\psi_{nm}, \tag{2.21}\]

This is still solvable using a matrix eigenvalue equation; for a spatial grid of size $N \times N$ the $N^2$ unknowns are found using a matrix sized $N^2 \times N^2$.

### 2.1.4 Nonparabolicity: Three-band Kane bandstructure model

The approximation made in the derivation of the effective-mass model is that the eigenstate has primarily the Bloch function of one single band, while the other bands influence only perturbatively. This is entirely true at the conduction band edge and remains a valid approximation for some higher energies; however, corrections must be made once the kinetic energies are a significant fraction of the bandgap. This is especially important in mid-IR QCLs where the energy scale is larger than in THz QCLs. This effect is known as “nonparabolicity,” reflecting the fact that the band dispersion in a bulk semiconductor is increasingly nonparabolic as $k$ moves away from the $\Gamma$-point.

A nice way of incorporating this into bandstructure calculations in bulk semiconductors was introduced long ago by Kane [101]. The approach is to directly diagonalize the Hamiltonian for a subset of bands, while still adding the effect of the rest perturbatively. When the conduction-like states are of interest, this might sometimes include eight bands (conduction, light holes, heavy holes, split-off holes $\times 2$ for spin), but if we assume spin degeneracy (weak spin-orbit coupling), the problem is reduced to four. Furthermore, the heavy-holes are decoupled from the rest, leaving three bands [104]. The three-band problem can in fact be even further reduced to a two-band one using a fictitious effective valence band [105], but for the purposes of this work the three-band problem is hardly any more complicated than the two-band one and has its advantages, for instance in the calculation of tunnel couplings (see Sec. ??).

We restrict our analysis here to the case of quantum wells; a more general method for quantum dots and using a full eight-band Hamiltonian was put forward in [106]. In choosing a quantum well system, our problem is reduced to only one dimension, under the same
assumption made in effective mass treatment that the in-plane directions are decoupled.

The eigenvalue equation used to calculate the growth-direction envelope functions in the
three bands ($C$, $L$, $S$ for conduction, light-hole, split-off) is [104]:

$$
\begin{bmatrix}
E_c & \sqrt{2}\alpha & -\alpha \\
\sqrt{2}\alpha^* & E_{lh} & 0 \\
-\alpha^* & 0 & E_{so}
\end{bmatrix}
\begin{bmatrix}
C \\
L \\
S
\end{bmatrix}
= E
\begin{bmatrix}
C \\
L \\
S
\end{bmatrix}.
$$

(2.22)

$E_c$, $E_{lh}$, and $E_{so}$ are the conduction, light-hole and split-off band edges (functions of $z$).

The operator $\alpha = ipz\sqrt{\frac{E_p}{6m_0}}$, where order is important as the Kane energy $E_p$ is a material
property thus having position-dependence. Evaluating the second and third rows of the
above equations gives the light-hole and split-off components in terms of the conduction
component:

$$
L = -\hbar \sqrt{\frac{E_p}{3m_0}} (E - E_{lh})^{-1} \frac{\partial}{\partial z} C
$$

(2.23)

$$
S = +\hbar \sqrt{\frac{E_p}{6m_0}} (E - E_{so})^{-1} \frac{\partial}{\partial z} C.
$$

(2.24)

Substituting these into the top row equation of (2.22) gives a Schrödinger-type equation
with energy-dependent effective mass for $C$ alone:

$$
-\frac{\hbar^2}{2m_0} \frac{\partial^2}{\partial z^2} \left[\frac{2}{3} \frac{E_p}{E - E_{lh}} + \frac{1}{3} \frac{E_p}{E - E_{so}}\right] \frac{\partial}{\partial z} C + E_c C = EC,
$$

(2.25)

with the effective mass $m^*(E, z)$ identified as:

$$
\frac{m_0}{m^*(E, z)} = \frac{2}{3} \frac{E_p}{E - E_{lh}} + \frac{1}{3} \frac{E_p}{E - E_{so}}.
$$

(2.26)

Although often ignored, the solutions $C$ from (2.25) are not in general orthogonal to each
other because of the energy dependence in the effective mass. This can, however, be resolved
if it is kept in mind that the full eigenstate also includes the $L$ and $S$ components; the
orthogonality relation is actually that $\langle C_1 | C_2 \rangle + \langle L_1 | L_2 \rangle + \langle S_1 | S_2 \rangle = 0$. True normalization must also include all three components as well: $\langle C | C \rangle + \langle L | L \rangle + \langle S | S \rangle = 1$.

The energy-dependence of the effective mass prevents a fully direct numerical solution of (2.25), so other methods have to be used. A common way to do this is using a shooting method [107, 102], which involves iterations, although a way of solving it directly under a third-order approximation of the $m^*$ energy dependence has been outlined as well [108]. Once the solutions $C$ are obtained, $L$ and $S$ can be calculated easily using (2.23) and (2.24).

Finally, in the spirit of the approximation of position-independent effective mass for the in-plane Schrödinger equation, the solutions are again simply plane waves. The effective mass used for the paraboloidal dispersion is somewhat difficult to define, so for the purposes of this thesis to simplify the analysis it is taken to be simply again the well effective mass; this approximation clearly only affects the subband dispersions, whereas the positions of the subband energy minima are more fully calculated, giving a reasonable description of the actual density of states.

Figure 2.2 shows a comparison of the conduction-band wavefunction components obtained in an effective mass versus a three-band Kane model for two different scenarios: a thin 10 nm well in the $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ material system where the lowest transition is in the mid-IR, and a thicker 30 nm well in $\text{GaAs/Al}_{0.15}\text{Ga}_{0.85}\text{As}$ where the lowest transition is in
the THz. It is clear that in mid-IR QCLs where kinetic energies reach to 100s of meV, the treatment of nonparabolicity is absolutely essential, whereas in THz QCLs an effective mass model is normally quite accurate, withholding minor corrections in excited states near the top of the well.

2.2 Optical selection rules and oscillator strengths

An optical transition is possible between states that have a nonzero dipole matrix element \( \vec{r}_{ab} = \langle \Psi_a | \vec{r} | \Psi_b \rangle \), a result that is visible directly from the optical Hamiltonian \( H_{opt} = q \vec{E}(\omega) \cdot \vec{r} e^{i \omega t} + c.c. \). For stimulated emission and absorption, this matrix element can vary or even disappear depending on the electric field polarization, whereas the spontaneous emission is inherently the interaction with all possible modes and therefore field directions.

The separation of variables assumption leads directly to strict selection rules on the allowed optical transitions, for both spontaneous and stimulated processes. For states separable as stated in (2.14) into \( \Psi = \psi_z(z) \psi_\parallel(x,y) \), the dipole matrix element for the pair \( a \) and \( b \) is:

\[
\langle \Psi_a | \vec{r} | \Psi_b \rangle = \hat{x} \langle \psi_{z,a} | \psi_{z,b} \rangle \langle \psi_{\parallel,a} | x | \psi_{\parallel,b} \rangle + \hat{y} \langle \psi_{z,a} | \psi_{z,b} \rangle \langle \psi_{\parallel,a} | y | \psi_{\parallel,b} \rangle + \hat{z} \langle \psi_{z,a} | z | \psi_{z,b} \rangle \langle \psi_{\parallel,a} | \psi_{\parallel,b} \rangle.
\]

(2.27)

Because the respective sets of \( \psi_z \) and \( \psi_\parallel \) are orthogonal, the dipole matrix element vanishes in the in-plane directions for a transition in the growth direction, and vice versa (in the growth direction for an in-plane transition). Therefore, transitions in both wavefunctions components will have no dipole matrix element in any direction. In quantum wells, this is known as the “intersubband selection rule,” although it holds as well for quantum dots if the same separation of variables is valid. (2.27) also has further consequences for stimulated processes, which is that transitions in \( \psi_z \) will only interact with the \( z \)-component of the electric field, and the same for the in-plane directions. In QCLs where it is the \( \psi_z \) that are intentionally engineered, this has the far-reaching ramification that intersubband transitions
only interact with light polarized in the growth direction.

2.2.1 Quantum wells

Since a quantum well heterostructure is intended for interaction with z-polarized radiation, the phrase “dipole matrix element” is commonly used to refer to this matrix element $z_{ab}$ specifically rather than the full $\vec{r}_{ab}$ in (2.27):

$$z_{ab} = \langle \psi_{z,a} | z | \psi_{z,b} \rangle \delta_{\vec{k}_a,\vec{k}_b}. \quad (2.28)$$

The requirement imposed by the Kronecker delta function reflects the in-plane selection rule and can also be thought of as in-plane momentum conservation, approximating the imparted photon momentum as zero. It is useful to define a term known as “oscillator strength,” which is the ratio of a transition’s spontaneous emission rate to the energy decay rate of a classical electron oscillator having the same energy: $f_{ab} = 2m_0E_{ba}|z_{ab}|^2/\hbar^2$. The oscillator strength obeys the sum rule that $\sum_{b \neq a} f_{ab} = 1$ for a sum over $b$ covering the entire Hilbert space of possible states (all bands). A more useful sum is over only intraband transitions, which under the assumption of parabolicity is simply that $\sum_{b \in \text{band} \neq a} f_{if} = m_0/m^* \ [104]$. Therefore, the oscillator strength defined above is often multiplied by the factor $m^*/m_0$ to give a “scaled oscillator strength” $f'$ that obeys the sum rule $\sum_{b \in \text{band} \neq a} f'_{ab} = 1$:

$$f'_{ab} = \frac{2m^*E_{ba}|z_{ab}|^2}{\hbar^2}. \quad (2.29)$$

$f'$ in this form is then a useful metric to assess the strength of optical transitions; a number that is a significant fraction of 1 indicates high optical transition strength, with the caveat that transitions from excited states can actually have oscillator strengths greater than 1 because of the negative sign in $f'$ for downward transitions.

In the in-plane directions, interaction with the optical field is usually treated using a classical theory such as the Drude model. However, the sum rule still gives a good way of understanding the finite absorption that arises from the continuum of states, and in
fact it was for this reason that the $m_0/m^*$ sum rule was first discovered by Sommerfeld and Bethe. The spatial-dependence of the effective mass and the nonparabolicity introduce some ambiguity into the definition of $f'$, but for the purpose of estimating optical transition strength this is not problematic. A more complete description accounting for this is found in [104].

2.2.2 Quantum dots

In quantum dots, under the separation of variables assumption, the analysis of dipole matrix elements in the growth direction can follow in the same way as for quantum wells, except that the in-plane momentum conservation must remain more generally a conservation of the in-plane state:

$$z_{ab} = \langle \psi_{z,a} | z | \psi_{z,b} \rangle \langle \psi_{\parallel,a} | \psi_{\parallel,b} \rangle.$$  \hspace{1cm} (2.30)

Oscillator strength and the sum rule for the z-direction also follow in the same way as for quantum wells. Regarding the in-plane directions, the oscillator strength concept can be extended to two (and three) dimensions [103]:

$$f_{ab}^{2D} = \frac{m^*(E_{ba})}{\hbar^2} (|x_{ab}|^2 + |y_{ab}|^2)$$  \hspace{1cm} (2.31)

$$f_{ab}^{3D} = \frac{2m^*(E_{ba})}{3\hbar^2} (|x_{ab}|^2 + |y_{ab}|^2 + |z_{ab}|^2).$$  \hspace{1cm} (2.32)

Oscillator strengths defined this way all obey the same sum rule, that $\sum_{b \neq a} f_{ab}^{1D} + \sum_{b \neq a} f_{ab}^{2D} + \sum_{b \neq a} f_{ab}^{3D} = 1$ (for sums only inside the conduction band). An interesting selection rule for the in-plane direction is that for a cylindrically symmetric potential where the angular momentum quantum number $m$ is a “good quantum number” ($[H, p_z] = 0$), in-plane transitions can only occur across transitions where the angular momentum differs by exactly $\hbar$, or in other words only when $|m_a - m_b| = 1$. This is in fact a simple statement of conservation of angular
momentum, since the photon is spin 1 and so must change the electron angular momentum by $\pm \hbar$ in the course of being created or destroyed. It is interesting that an aspect of the electric field quantization is present despite the field being treated classically.

### 2.2.3 Nonparabolicity

For energy eigenstates calculated using the 3-band Kane model of Sec. 2.1.4, the dipole matrix element should be evaluated as:

$$z_{ab} = \langle C_a | z | C_b \rangle + \langle L_a | z | L_b \rangle + \langle S_a | z | S_b \rangle,$$

which comes from the assumption that the matrix elements between the Bloch functions $u_C$, $u_L$, $u_S$ are negligible. Sometimes $z_{ab}$ is calculated from only $C$ (normalized by itself), but this tends to be an overestimation because the $L$ and $S$ components usually contribute little, partly because they have opposite sign (see 2.23, 2.24).

A modified sum rule under the effective two-band Kane model is derived in [104], but is not necessary if the oscillator strength is merely meant to be a rough measure.

### 2.3 Overview of existing QCL transport models

Over the years, a wide range of techniques have been used to model electron transport, gain, and optical nonlinearity in QCL active regions. An overview will be given here in order of quantum mechanical complexity, which loosely follows the overall complexity. Further review and discussion can be found in [102, 109].

The simplest QCL transport models are the rate equation models, which descend from the basic rate equations used since the invention of the laser [110, 111, 112, 113, 114]. These models include only population transfer between the different subbands in the structure; therefore some assumptions must be made about the electron distributions within subbands and the rates are averaged over the distributions, or the rates are simply estimated based on experimental or other data. Even though rate equation models are simple and include only
a basic description of the physics, they have contributed much to the field by introducing simple analytic equations to make clear the dependence of performance metrics on certain quantities (for example the dependence of gain on a specific transition rate). In addition, it is only these simplest models that can be coupled with simulations of the mode dynamics, to capture more global phenomena such as spatial hole burning, mode competition, and mode-locking [115, 116, 117, 118].

The rate equation concept can also be extended to include the in-plane electron motion, although this is at the expense of losing analytical simplicity. This method is called the “Ensemble Monte Carlo” (EMC) method, and has been used widely for the design and optimization of both mid-IR and THz QCLs [119, 120, 121, 122, 123, 124, 125, 126]. Over the years, the EMC approach has grown to include higher levels of detail, with notable milestones being the inclusion of electron-electron scattering [120] and the influence of a strong optical field [124]. The EMC method is also numerically light enough that it can be used self-consistently, for example in conjunction with Poisson’s Equation to account for the effects of space charge.

The rate equation and EMC approaches are categorized as “semiclassical,” because while the bandstructure is calculated from a quantum theory, the transport occurs only by “hopping” transitions between state populations, described with the transition rates obtained from Fermi’s Golden Rule. This treatment neglects any effects of correlation between the heterostructure eigenstates, because the state populations are the extent to which the electrons are described. Correlations can become important if scattering tends to push electrons into states that are not energy eigenstates of the heterostructure, which is now well-known to occur in QCLs, particularly at thick tunneling barriers. Sometimes the effects of tunneling are approximated in the rate equation models [127], but more sophisticated theories are often required, for which the field has turned to the density matrix (DM) and Non-equilibrium Green’s Function (NEGF) formalisms.

The DM formalism was actually used in the very first theoretical conception of a QCL by Kazarinov and Suris in 1971 [128]. Density matrices allow for correlations between states, which leads to a nice description of tunneling as will be described in the following sections.
Given equations of motion for the density matrix, the steady-state can be solved analytically for simpler special cases [129, 130], or more generally using a numerical solution [131, 132, 133]. The density matrix problem is sometimes also solved temporally for subband-averaged quantities [134], and more recently resolved in the in-plane dynamics [195, 196]. Description of density matrix transport modeling is kept short here as it is the subject of most of this thesis.

Finally, a still more detailed analysis of quantum transport in QCLs is possible through the use of non-equilibrium Green’s functions (NEGF). NEGF extends the machinery of many-body theory into the far-from-equilibrium regime, allowing quantum correlations in both space and time to be included and account for scattering effects to arbitrary order via the appropriate self-energies. In practice, the equations of motion that are solved in NEGF calculations require the self-consistent, iterative computation of quantities with non-local dependence on space, in-plane momentum, and energy, making the memory and CPU requirements extremely expensive. NEGF models have made enormous contributions to the field of QCLs [137, 138, 139, 140], but thorough review is beyond the scope of this thesis.

2.4 Density matrix transport modeling

This section outlines the basic density matrix theory used in this thesis to model transport in QCLs. The general approach is to calculate a full density matrix equation of motion, solve it in the steady-state, and from this extract properties such as current, optical gain, and optical nonlinearity. The optical field is accounted for coherently and nonperturbatively by allowing the Hamiltonian and density matrix to rotate at certain frequencies, and allowing these frequencies to mix.

2.4.1 The density matrix

The density matrix provides a way of representing a mixed state, which is a statistical description allowing for varying degrees of coherence. As an illustrative example, we consider an electron with two possible states: \( \phi_a \) and \( \phi_b \). On one hand, the electron state \( \Psi \) could
be a superposition, such as $\Psi = (\phi_a + \phi_b)/\sqrt{2}$. In writing down $\Psi$ (which represents a pure state), a coherence had to be assigned, which was in this case chosen as $c_a^*c_b = 1/2$. On the other hand, if this information is lacking, we cannot describe the electronic state as a superposition, since we have no coherence to write down. The density matrix $\rho$ provides a way to represent this lack of information:

$$\rho = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \rightarrow \text{completely coherent superposition}$$

$$\rho = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \text{completely incoherent superposition} \quad (2.34)$$

The elements of the density matrix are formally $\rho_{ab} = \langle c_a^*c_b \rangle$, with $\langle \rangle$ denoting either an expectation value or ensemble average and $c_{a,b}$ being the state amplitudes in the basis states. Therefore, the diagonals of the density matrix $|c_a|^2$ are the populations, and the off-diagonals are coherences. The magnitude of the coherence $\rho_{ab}$ for $a \neq b$ can be anywhere from 0 to $\sqrt{n_an_b}$ ($n_{a,b}$ are the populations), these being the limits of zero and complete existence of the coherence information. The populations themselves are normalized by the condition $Tr(\rho) = 1$.

The equation of motion for the density matrix of an isolated quantum system is the Liouville-von Neumann equation:

$$\dot{\rho} = -\frac{i}{\hbar}[H,\rho] = \mathcal{L}\rho. \quad (2.35)$$

$\mathcal{L} = -\frac{i}{\hbar}[H,\ldots]$ is a superoperator (known as the “Liouville superoperator”), since it acts in a general sense on $\rho$ that is not describable using simply matrix multiplication. It can be shown that the influence on any matrix element $\rho_{ab}$ from any other element $\rho_{cd}$ is:

$$\mathcal{L}_{ab,cd} = -\frac{i}{\hbar}(\delta_{bd}H_{ac} - \delta_{ac}H_{db}), \quad (2.36)$$
a form that is useful later when steady-state equations for the density matrix are developed.

However, truly isolated quantum systems tend not to exist in reality; there is always some interaction with an environment. Additionally, the environment is very large in comparison to the system of interest, and so can only ever be treated statistically. Under the assumption that the environment acts perturbatively on the system, this gives rise to “incoherent” evolution on top of the “coherent” Liouville-von Neumann evolution above. Fermi’s Golden Rule transitions are one familiar instance of incoherent evolution, which do not carry coherence information but only act to change the populations, but in density matrix models this encompasses an even broader set of effects, such as decoherence, which is the decay of coherence information. There can in fact be, even more generally, incoherent influence of any DM element on any other, a concept explored in Chapter 6, but in many cases the transitions and decoherence are sufficient in describing the physics of interest.

2.4.2 Tunneling in density matrix models

The common drawback to the purely semiclassical QCL transport models is that the true physics of resonant tunneling is not properly accounted for. Resonant tunneling is almost universally employed in QCL active regions, usually for injection of electrons into an upper lasing subband, or extraction from a lower.

The problem with a semiclassical theory can be understood as follows. If the subband energy structure is solved for across a resonant tunneling barrier, the result is that two anticrossed states appear, which both extend significantly to both sides of the barrier. This is illustrated in Figure 2.3 as the “delocalized” basis. If a semiclassical transport theory is applied given this bandstructure, transitions to either of these states will effectively move electrons instantly across the barrier, independently of the actual barrier width $L_b$. This causes a semiclassical theory to predict the current at resonance to have almost no dependence on $L_b$, which is widely known not to be the case: current at resonance tends to decrease with $L_b$. The physical reason is that scattering tends to localize electrons as they cascade down the structure, so that they are actually first trapped upstream of the tunneling barrier.
Density matrix models allow for a nice way to mimic this kind of effect by making a “localized” basis choice, also illustrated in Figure 2.3. This is done by solving the Schrödinger equation separately on each side of the barrier, to yield two distinct basis states on each side. Since these states are not energy eigenstates of the full heterostructure Hamiltonian, they share an off-diagonal matrix element in $H$, known as the tunnel coupling, which drives the tunneling process. Transitions and dephasing are then calculated using this basis, so that the effect of scattering to a localized state is forced into the model. It is important to note that it is actually this incoherent evolution that is basis-dependent, whereas the coherent evolution in (2.35) is not.

Another option to properly account for tunneling is to use a fully generalized scattering superoperator, in which any density matrix element can influence any other; such a theory can work using the less ambiguous delocalized basis [195, 196]. Scattering by localization appears naturally in that coherence gets transferred to the delocalized basis with a phase such that electrons sit behind the barrier. Although it is not always necessary to go to this great length in QCL transport modeling, there are certain active regions where it is needed because an obvious localized basis choice does not exist.

Figure 2.3: Basis choices around a resonant tunneling barrier.
2.4.3 Scattering in density matrix models

The Liouville-von Neumann equation is the complete equation of motion for a closed quantum system. In a QCL, this means that if we truly knew everything about the total system (the heterostructure profile and its roughness, the location of all impurities, the state of all phonon modes, everything about the environment surrounding the laser, etc.), in principle the density matrix of the total system could be computed at any later time. However, we do not possess this level of information, nor would it be a remotely tractable problem even if we did. For this reason, much of the influence of the environment must be treated statistically, and it is actually this lack of information that requires a density matrix, as opposed to pure state, treatment in the first place. The density matrix that we speak of represents only a subsystem coupled to a much larger environment, which implies a loss of phase information to the environment, and thus incoherent evolution of the subsytem. This is at the root of the enormous topics of open quantum systems and decoherence theory, about which further reading is found in [141] and [142].

The most familiar type of incoherent evolution is the Fermi’s Golden Rule transition. The transition rate for an elastic process between state $a$ and a continuum of final states $[b]$ is given by:

$$W_{a\rightarrow[b]} = \frac{2\pi}{\hbar} |V_{ab}|^2 g_b(E_a),$$  \hspace{1cm} (2.37)

where $g_b(E_a)$ is the density of final states $b$ at the energy $E_a$. These are the states $b$ for which energy conservation is satisfied, so $V_{ab}$ is the corresponding interaction Hamiltonian matrix element. The bare energy structure, which is the coherent evolution, enters into Fermi’s Golden Rule through the density of states, whereas the interaction is treated perturbatively on top of this; the approximation giving rise to irreversible incoherent evolution is that any action by $V_{ab}$ is interrupted by action of the coherent Hamiltonian on a much shorter time scale [143]. Therefore, by assuming some evolution to be incoherent we are inherently assuming it to act perturbatively. This assumption helps greatly to allow a treatment of statistical effects from the environment, but there are of course cases where it might break
down, such as in the presence of extreme statistical interface roughness. In addition, this approximation breaks down under strong coupling with the environment, in which case either (a) memory effects in the subsystem must be accounted for, or (b) the part of the environment that is strongly-coupled must be included as part of the system. In Chapters 3 and 4, approach (b) is taken to account for the effects of strong-coupling of electrons with phonons in quantum dots.

Throughout physics, dissipation in open quantum systems is often classified into “T₁” and “T₂” times. T₁ times describe transitions, or in other words the movement among populations, whereas T₂ times describe dephasing, the decay of coherences that either accompanies transitions (lifetime dephasing) or happens alone (pure dephasing). There are in fact, even more generally, “T₃” times which will be explored in Chapter 6.

A transition between two states a and b at time $\tau_{ab}$ affects the density matrix in multiple places, specifically that:

$$
\dot{\rho}_{aa} = -\frac{1}{\tau_{ab}} \rho_{aa} \quad \dot{\rho}_{bb} = +\frac{1}{\tau_{ab}} \rho_{aa} \quad \dot{\rho}_{ab} = -\frac{1}{2\tau_{ab}} \rho_{ab} \quad \dot{\rho}_{ba} = -\frac{1}{2\tau_{ab}} \rho_{ba}.
$$

(2.38)

The first two equations describe the actual population transfer, whereas the second two are the accompanying lifetime dephasing. Although it is possible to find a superoperator to act on the density matrix in this way, a convenient way of doing this is to introduce the Lindblad form [144]:

$$
\dot{\rho}|_X = L_X \rho = C_X \rho C_X^\dagger - \frac{1}{2} \left( C_X^\dagger C_X \rho + \rho C_X^\dagger C_X \right).
$$

(2.39)

The $C_X$ are “collapse” or “jump” operators, and evolution in the Lindblad form has the significance that it is guarantees trace preservation and complete positiveness in the density matrix, for any $C_X$. $X$ labels some specific process, and for the generic transition $a \rightarrow b$ described above, the associated jump operator is simply $C_{a \rightarrow b} = \frac{1}{\sqrt{\tau_{ab}}} |b\rangle \langle a|$. For the full group of transitions $X$ that act on the system in a sense uncorrelated to each other, the full
system evolution is a sum of the Lindblad evolutions: \( \dot{\rho}_{\text{all transitions}} = \sum X L_X \rho \).

The pure dephasing times, known as \( T^{\ast}_2 \), are in principle describable using the Lindblad form, since pure decoherence certainly obeys trace preservation and complete positiveness, but it is generally simpler to include this as just a single element in the evolution superoperator: if \( D \) is this superoperator for pure dephasing, then the pure decoherence of \( \rho_{ab} \) with time \( T_{2,ab} \) is incorporated by setting \( D_{ab,ab} = -1/T_{2,ab} \).

### 2.4.4 Periodic boundary conditions

With only a few exceptions, QCL active region designs use a single “module” of only a few wells and barriers which is repeated a large number (often hundreds) of times. An example is illustrated in Figure 2.4, and this can be reflected in the following form of the Hamiltonian, dipole matrix, and density matrix:

\[
H = \begin{bmatrix}
\ddots & \vdots & \ddots \\
(H_0 - \Delta) & (H_1) & (0) \\
\cdots & (H_{-1}) & (H_0) & (H_1) & \cdots \\
(0) & (H_{-1}) & (H_0 + \Delta) \\
\ddots & \vdots & \ddots 
\end{bmatrix} \tag{2.40}
\]

\[
z = \begin{bmatrix}
\ddots & \vdots & \ddots \\
(z_0 - L) & (z_1) & (0) \\
\cdots & (z_{-1}) & (z_0) & (z_1) & \cdots \\
(0) & (z_{-1}) & (z_0 + L) \\
\ddots & \vdots & \ddots 
\end{bmatrix} \tag{2.41}
\]
Figure 2.4: An example of periodicity in a QCL active region, with interactions occurring both inside and between modules.

\[
\rho = \begin{bmatrix}
  \cdots & \cdots & \cdots \\
  (\rho_0) & (\rho_1) & (0) \\
  \cdots & (\rho_{-1}) & (\rho_0) & (\rho_1) & \cdots \\
  (0) & (\rho_{-1}) & (\rho_0) \\
  \cdots & \cdots & \cdots 
\end{bmatrix}
\]

(2.42)

Terms inside the () are submatrices which have the size \( N \times N \), \( N \) being the number of states in each module. The submatrices with subscript 0 are intramodule operators, and subscript \( \pm 1 \) denotes intermodule operators between next-neighbor modules. Operators beyond next-neighbor are ignored, approximating the coherence length as well shorter than one module, although the formalism presented easily extends to arbitrarily-further interactions. \( \Delta \) and \( L \) hold, respectively, the progressive differences in energy and position between neighboring modules, thus influencing only the diagonal elements. \( \Delta = E_{\text{mod}} I_N, \) where the neighboring-module energy separation \( E_{\text{mod}} \) will vary with the applied electric field, whereas \( L = L_{\text{mod}} I_N, \) where the neighboring-module spatial separation \( L_{\text{mod}} \) is fixed for a given heterostructure design. No progressive difference is applied to \( \rho \), forcing the state of the electrons to be periodic. In reality this might differ slightly from the truth, for instance due
to nonuniformities in the optical mode across the active region, which has strong influence
on the electrons, or due to other more global effects such as electric field domain formation.
However, this approximation is useful here for the sake of getting insight into the operation
of the active region alone, without effects belonging more generally to the coupled active
region/cavity system.

Under periodic boundary conditions, the normalization condition can be expressed per
module: \( Tr(\rho_0) = 1 \). There is no normalization condition on \( \rho_{\pm 1} \).

### 2.4.5 Harmonicity

The optical field itself is a time-varying harmonic Hamiltonian:

\[
H_{opt} = q F z e^{i\omega t} + c.c. , \tag{2.43}
\]

where \( z \) is the full position operator and \( F \) is the optical field. To generally allow for both
dc and optical components in the Hamiltonian, we write it as:

\[
H(t) = H^{(-\omega)} e^{-i\omega t} + H^{(0)} + H^{(\omega)} e^{i\omega t}, \tag{2.44}
\]

and the density matrix could be expected to respond at (for now) these same frequencies:

\[
\rho(t) = \rho^{(-\omega)} e^{-i\omega t} + \rho^{(0)} + \rho^{(\omega)} e^{i\omega t}. \tag{2.45}
\]

As will be explored in Chapter 5, \( H \) and \( \rho \) can be even further generalized so that
they include rotation at any set of arbitrary frequencies \( \omega_m \), which is used for the study of
nonlinear optics:

\[
H(t) = \sum_m H^{(\omega_m)} e^{i\omega_m t} \tag{2.46}
\]
\[
\rho(t) = \sum_m \rho^{(\omega_m)} e^{i\omega_m t}.
\]

(2.47)

Mixing between \( H \) and \( \rho \) occurs by the \([H, \rho]\) commutator in the Liouville-von Neumann equation. Ordinarily, the calculation of optical properties involves a “rotating wave approximation” (RWA), in which specific elements in \( H \) and \( \rho \) are assumed to rotate in only one direction based on knowledge of the state energies (the energy-conserving direction), but for the purposes of this work such an approximation is not made, as it takes more effort to make the RWA than not.

### 2.4.6 Steady-state solution

If \( N_f \) is the number of optical frequencies included, then the total number of terms in 2.46 and 2.47 is \( 2N_f + 1 \), to include both the positive and negative frequencies and the dc. Given three submatrices per unit cell, each of size \( N^2 \) (see 2.42), there are then \( 3(2N_f + 1)N^2 \) unknowns. There are various ways to solve for the steady-state, given in some detail in Appendix A, but the general concept is that there is a large matrix that relates any possible unknown to any other to form a linear system of equations, which is solveable once a normalization condition is applied to the dc populations. For the coherent part of the evolution, applying the Liouville-von Neumann equation with 2.40 and 2.42 leads to the coherent evolution of any particular submatrix:

\[
\dot{\rho}_p|_{coh} = \sum_q [H_{p-q}, \rho_q] - p\Delta\rho_p.
\]

(2.48)

Next, once these are expanded into their steady-state harmonics, we have the conditions also relating the frequency components:

\[
i\omega_m\rho^{(\omega_m)}_p = \sum_{qn} \left( -\frac{i}{\hbar} \left[ H^{(\omega_m - \omega_n)}_{p-q}, \rho^{(\omega_n)}_q \right] - \delta_{pq}q\Delta^{(\omega_m - \omega_n)}\rho^{(\omega_n)}_q \right) + \text{incoherent}.
\]

(2.49)
This equation is quite insightful, as we see that different frequency components in \( \rho \) are coupled to each other through their difference frequency in \( H \); this is the frequency mixing effects at work. Frequency mixing occurs even for the case of only one optical frequency as in 2.44, 2.45, for instance in that \( H(\omega) \) can couple \( \rho(-\omega) \) to \( \rho(0) \) or \( \rho(0) \) to \( \rho(\omega) \).

It is worth noting that the harmonic time dependence is also included in \( \Delta \), the energy separation between neighboring modules. This subtle detail cannot be ignored in some cases, in particular for the study of terahertz difference frequency generation.

Further derivation, including the incoherent contribution, will be given in Appendix A.

2.4.7 Extraction of velocity

Once the steady-state density matrix is found, we can extract the expectation value of velocity at different frequencies \( \langle v(\omega_m) \rangle \), leading to the transport properties of interest, for example:

\[
J = N_d q \langle v(0) \rangle \rightarrow \text{Current density} \tag{2.50}
\]

\[
\chi^{(1)}(\omega) = \frac{N_d q \langle v(\omega) \rangle}{i \omega \epsilon_0 F(\omega)} \rightarrow \text{First-order optical susceptibility} \tag{2.51}
\]

\[
\chi^{(2)}(\omega_1 = \omega_3 - \omega_2) = \frac{N_d q \langle v(\omega_1) \rangle}{i \omega \epsilon_0 F(\omega_2) F(\omega_3)} \rightarrow \text{Difference-frequency susceptibility (second-order)} \tag{2.52}
\]

It is tempting to use the position instead of the velocity in evaluating the optical properties, but issues can arise with this approach because of the infinite periodic nature of our system in which polarization per unit cell is not uniquely defined.

To get the velocity, we require knowledge of the velocity operator, which for the coherent part of the evolution is expressible as \( v_{coh} = \frac{i}{\hbar} [H,z] \). \( H \) has different frequency components, making it appear that the velocity operator will as well; however, because the ac \( H \) components are proportional to \( z \), the ac velocity operator components vanish and we are left
with only $v_{coh}^{(0)}$. From here forward, this will be referred to as $v$; the incoherent parts will be addressed in Appendix A.

$v$ will have periodic boundary conditions similar to 2.42, in which there is no progressive difference needed along the diagonal. By evaluating $v = \frac{i}{\hbar}[H, z]$ using 2.40 and 2.41, we arrive at the submatrices of $v$:

\[-i\hbar v_{-1} = [H_{-1}, z_0] + [H_0, z_{-1}] + \Delta z_{-1} - LH_0\]
\[-i\hbar v_0 = [H_{-1}, z_1] + [H_0, z_0] + [H_1, z_{-1}]\]
\[-i\hbar v_1 = [H_0, z_1] + [H_1, z_0] - \Delta z_1 + LH_1.\] (2.53)

The velocity is then evaluated as $\langle v \rangle = \text{Tr}(v\rho) = \text{Tr}(v_{-1}\rho_1 + v_0\rho_0 + v_1\rho_{-1})$, where the trace runs over one module, at the different frequencies. A way of extracting the incoherent velocity is given in Appendix A.

A more general method to extract the velocity is by use of the time evolution superoperator itself, which would include both the coherent and incoherent evolution. If this superoperator is $X$, then the expectation value for velocity is:

$$\langle v \rangle = \text{Tr}(z\dot{\rho}) = \text{Tr}(zX\rho) = \sum_{ab,cd} z_{ab}X_{ab,cd}\rho_{cd}.\] (2.54)

By regrouping terms, the velocity operator can be identified since $\langle v \rangle = \text{Tr}(v\rho) = \sum_{cd} v_{dc}\rho_{cd}$:

$$v_{dc} = \sum_{ab} z_{ab}X_{ab,cd} \rightarrow v_{cd} = \sum_{ab} z_{ab}X_{ab,cd}^*.\] (2.55)

Rearranging variables again, the key distinction between solving for steady-state and extracting velocity is visible:
\[ \dot{\rho}_{ab} = \sum_{cd} X_{ab,cd} \rho_{cd} = 0 \quad \rightarrow \quad \text{Steady-state equation} \]

\[ v_{ab} = \sum_{cd} X^*_{cd,ab} z_{cd} \quad \rightarrow \quad \text{Velocity operator.} \quad (2.56) \]

The important difference between the two is that the subscripts of \( X \) are switched: for steady-state analysis we are concerned with evolution moving \emph{into} the element \( ab \), whereas to extract velocity we are interested in evolution \emph{out of} \( ab \). In an infinite system this difference sets rules on how 2.54 should be evaluated, which are outlined in Chapters 5 and 6 where this general approach is taken.
CHAPTER 3

Quantum Dot Cascade Lasers

3.1 Background

The motivation for a quantum dot cascade laser (QDCL) lies expressly in the discrete density of states, which is expected to cause a suppression of LO-phonon emission across a lasing transition. In an ordinary QCL, LO-phonon emission is possible between subbands spaced by energies different from the LO-phonon energy ($E_{LO}$) because of the continuous spectrum of energies above the subband minimum and the momentum carried by the LO-phonons (see Fig. 1.4). If the density of states is discrete, however, it is easy to imagine that LO-phonon emission might be avoided between sublevels not spaced by $E_{LO}$. This effect is known as “phonon bottleneck,” and has been studied now for quite some time, not limited to the context of QCLs [87, 86, 145, 146, 147, 148, 149, 150, 151].

In a view of the electron-LO-phonon interaction as a scattering mechanism, it would seem that energy relaxation is only possible across transitions separated by exactly $E_{LO}$, at least within some broadening width. However, the reality is that the discrete density of electronic states combined with a discrete LO-phonon energy causes a strong-coupling of the two to occur. The resulting long-lasting strongly-coupled electron-phonon quasiparticles are known as “polarons”, and their existence causes a quite new transport physics to emerge, including a qualitative difference in the energy selectivity for relaxation. A significantly reduced dephasing rate as compared to conventional QCLs is also expected to have profound implications for transport.

Previous theoretical studies on quantum dot cascade lasers and quantum dot superlattices have been performed using the NEGF formalism, for lens-shaped self-assembled quantum
dots in the high-confinement regime [65, 66] and also for a nanopillar geometry where the diameter was allowed to vary from the quantum to classical regime [152, 153]. In these works, the electron-phonon interaction is accounted for by a phonon Green’s function which enters into the self-energy; it is therefore represented as an average field which is assumed to remain at thermal equilibrium. The phonon decay, which broadens the interaction, is treated by introducing an anharmonicity in the phonon Green’s function.

Attempting to model a QDCL using the electron density matrix runs into the issue that no obvious density of states exists to be used to describe Fermi’s Golden Rule transitions (see 2.3). On the physical level, the issue is that irreversibility is no longer a good approximation, owing to the fact that the system (electron) is not weakly-coupled to its environment (LO-phonons), which is the common assumption made to justify incoherent evolution in the first place (see Section 2.4.3). Importantly, the electron system is now able to strongly influence the phonons, which can return to affect the electrons. This suggests that a proper model of transport should track both on equal footing.

This chapter outlines a density matrix model for QDCLs that accounts for the existence of polarons by using an electron-phonon tensor product Hilbert space, such that the phonons are brought in as part of the system. While the previous NEGF studies do account for strong electron-phonon coupling in some sense, the LO-phonons are assumed to remain in thermal equilibrium, thus playing the role of a system bath; here the LO-phonon mode states are allowed to move freely and the role of the system bath is pushed further to the acoustic phonons. Results predict sufficient gain for lasing at room temperature, but also significant new challenges arising from the polaron formation, which call for a new strategy suggested in Chapter 4.
3.2 Polaron physics

3.2.1 Basic model

The relaxation of electrons by interaction with an LO-phonon in quantum dots can be understood by the following dynamical picture. If we prepare the electron initially in the upper state of a two-level system tuned closed to $E_{LO}$, the electron can relax to the lower state, accompanied by emission of an LO-phonon. At this point, since the LO-phonon still exists on top of the relaxed electron, it can be re-absorbed, followed by re-emission, etc. If in principle there were no other interrupting effect, this cycle, a Rabi oscillation, would continue indefinitely. The polaron coupling strength which drives this oscillation tends to be on the order of a few meV, leading to THz-level oscillation frequency.

The most important interruption that occurs is the decay of the LO-phonon itself; there is a certain probability per time that the LO-phonon exists that it will decay into various pairs of acoustic phonons, on the timescale of several ps [151, 154] (approximately 2-7 ps depending on temperature). This part of the interaction is irreversible because the acoustic phonon modes do form a continuum. Ultimately this is the pathway by which the electron can relax irreversibly to the lower state, since once the phonon decays it will not be reabsorbed. Importantly, because the timescale of acoustic phonon decay is longer than the frequency of the Rabi oscillations, many oscillations can occur over the course of full relaxation.
Figure 3.1 shows a schematic and the time dynamic energy relaxation, for rough values of the parameters \( \Omega_{pol} \) (polaron coupling strength) and \( \tau_{\text{decay}} \) (LO-phonon decay time). The time dynamics are shown for transitions tuned to \( E_{LO} \) and detuned by 5 meV. Although overall energy relaxation is most efficient across transitions tuned to \( E_{LO} \), this mechanism is still significant for a wide range around \( E_{LO} \). It is also worth noting that the Rabi oscillation frequency speeds up for transitions detuned from \( E_{LO} \), which is a result of the energy eigenstates of the electron-phonon system being further apart.

### 3.2.2 Electron-LO phonon coupling

The electron-LO-phonon interaction is described by the Fröhlich Hamiltonian \( \hat{H}_f \), which includes all modes simultaneously. Assuming bulk plane-wave LO-phonons with wavevectors \( \vec{k} \), \( \hat{H}_f = \sum \hat{F}_{\vec{k}} \), where \( \hat{F}_{\vec{k}} \) is the Fröhlich Hamiltonian for single mode \( \vec{k} \), expressed as:

\[
\hat{F}_{\vec{k}} = \frac{A_k}{k\sqrt{V}} \left( e^{i\vec{k} \cdot \vec{r}_b} b_{\vec{k}} + e^{-i\vec{k} \cdot \vec{r}_b} b_{\vec{k}}^\dagger \right),
\]

where the constant \( A = \sqrt{\frac{E_{LO} q^2}{2}} \left( \frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_{dc}} \right) \). \( V \) is the crystal volume, \( b_{\vec{k}} \) and \( b_{\vec{k}}^\dagger \) are annihilation and creation operators, and \( \epsilon_\infty \) and \( \epsilon_{dc} \) are the high- and low-frequency bulk permittivities. For a particular transition which involves electronic states \( \psi_1 \) and \( \psi_2 \), we form product states with LO-phonon modes \( \vec{k} \) and define the matrix elements:

\[
F_{\vec{k},T} \equiv \langle \psi_1; 0 | \hat{H}_f | \psi_2; 1_{\vec{k}} \rangle = \langle \psi_1; 0 | \hat{F}_{\vec{k}} | \psi_2; 1_{\vec{k}} \rangle,
\]

where for example \( |\psi_1; 0\rangle \) describes an electron in state “1” with not existing phonon, and \( |\psi_2; 1_{\vec{k}}\rangle \) describes an electron in state “2” with one phonon in mode \( \vec{k} \). We now follow previous works and introduce a particular LO-phonon mode \( T \), which is a superposition of plane-wave modes, defined through any number state \( |n_T\rangle \)\(^{[146, 150]} \)
\[ |n_T\rangle \equiv \frac{1}{\sqrt{\sum_k |F_{k,T}|^2}} \sum_k F_{k,T}^* |n_k\rangle. \]  

(3.3)

Under assumption of LO-phonon degeneracy, this mode remains an energy eigenmode. This is valid given that the interaction strength falls off rapidly for phonon wavevectors not much larger than the inverse dot size (~ 20 nm), so that the relevant phonon modes comprise only a small part of the Brillouin zone close to the Γ-point. The coupling strength to mode \( T \) is then:

\[ \langle \psi_1; 0 | \hat{H}_F | \psi_2; 1_T \rangle = \sqrt{\sum_k |F_{k,T}|^2} \equiv \Omega_{\text{pol},T}, \]  

(3.4)

while it can be shown that the matrix element involving any orthogonal mode is zero. Therefore, the problem reduces to one involving only a single mode per electronic transition.

In the form of an integral over \( \vec{k} \), the expression for \( \Omega_{\text{pol},T} \) becomes:

\[ \Omega_{\text{pol},T}^2 = \frac{A^2}{(2\pi)^3} \int d^3\vec{k} |F^{(T)}(\vec{k})|^2 / k^2, \]  

(3.5)

where we have defined the form-factor for the transition

\[ F^{(T)}(\vec{k}) \equiv \langle \psi_1 | e^{i\vec{k} \cdot \vec{r}} | \psi_2 \rangle. \]  

(3.6)

The displacement field for such a mode can be constructed using (3.3). To obtain a more physical understanding of this mode, we consider the lowest energy intersublevel transition in a simple cylindrical quantum dot with a height of 30 nm and a diameter of 20 nm, with an infinite confinement potential on all sides. The wavefunctions are thus products of the infinite square well ground and first excited states in the axial direction with the circular well ground state in the cross-sectional plane. A plot of upward and radial displacements for the associated phonon mode are depicted in Figure 3.2. This helps to justify our use of an unbounded plane—wave basis - the results are not very different than if confined modes
Figure 3.2: The particular phonon mode interacting with a cylindrical QD lowest-lying transition: (left) upward and (right) radial displacements. Both are in separate arbitrary units, and the radial displacement is shown at a phase $\pi/2$ relative to that of the upward.

were used.

3.5 is a general expression for the polaron coupling between any two three-dimensional electron states. It is often advantageous to assume cylindrical symmetry (circular pillar cross-section) to simplify the calculations, to avoid having to do a six-dimensional integral (three spatial dimensions for the form factors and three dimensions in $\vec{k}$). Expressing 3.5 in cylindrical coordinates gives:

$$\Omega_{\text{pol},T}^2 = \frac{A^2}{(2\pi)^3} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \int_{0}^{2\pi} d\phi_k dk_z dk_{\parallel} \frac{|F^{(T)}(k_z, k_{\parallel}, \phi_k)|^2}{k_z^2 + k_{\parallel}^2}. \quad (3.7)$$

From here, we need to calculate the form-factors, now as a function of $\vec{k}$ in cylindrical coordinates. The wavefunctions come from separation of variables, where the in-plane wavefunctions are the solutions to the infinite circular well, and the z-direction wavefunctions are solutions of the heterostructure profile in the growth direction. These involve Bessel functions, and have definite angular momentum $m\hbar$:  

\[ 50 \]
\[ \psi_1 = Z_1(z)R_1(r)\frac{1}{\sqrt{2\pi}} e^{i m_1 \phi} \quad (3.8) \]
\[ \psi_2 = Z_2(z)R_2(r)\frac{1}{\sqrt{2\pi}} e^{i m_2 \phi} \quad (3.9) \]

The factor \( \sqrt{1/2\pi} \) has been inserted in the spirit of separate normalization of the three wavefunction components: the other two normalizations are \( \int dz Z^2 = 1 \) and \( \int_0^\infty drr^2 = 1 \).

Next, the form factor \( \mathcal{F}^{(T)}(\vec{k}) \) is calculated, substituting the wavefunctions \( \psi_1 \) and \( \psi_2 \) into its definition 3.6 and switching the integration to cylindrical coordinates:

\[ \mathcal{F}^{(T)}(\vec{k}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_0^{2\pi} \int_0^{\infty} d\phi dr dz r Z_1(z)Z_2(z)R_1(r)R_2(r) e^{i(m_2-m_1)\phi} e^{i\vec{k} \cdot \vec{r}} \quad (3.10) \]

It is very useful at this point to decompose the in-plane part of the plane-wave phonons into cylindrical waves:

\[ e^{i\vec{k} \cdot \vec{r}} = e^{ik_z z} \sum_{n=-\infty}^{+\infty} i^{-n} J_n(k_{||} r) e^{in(\phi - \phi \vec{k})}. \quad (3.11) \]

We can clearly see that the only term in this equation that must be kept is that in which \( n = m_1 - m_2 \). This result is very important, and comes from conservation of angular momentum: the phonon which couples to a transition has the angular momentum of that transition.

Substituting 3.11 into 3.10, we find that:

\[ \mathcal{F}^{(T)}(k_z, k_{||}, \phi \vec{k}) = i^{m_1-m_2} e^{-i(m_1-m_2)\phi \vec{k}} \mathcal{F}_z^{(T)}(k_z) \mathcal{F}_{||}^{(T)}(k_{||}) \],

where we have introduced the “partial form-factor”: 51
\[ F_{\parallel}^{(T)}(k_{\parallel}) \equiv \int_{0}^{+\infty} drr R_1(r) R_2(r) J_{m_1-m_2}(k_{\parallel}r) \] (3.13)

\[ F_{z}^{(T)}(k_z) \equiv \int_{-\infty}^{+\infty} dz Z_1(z) Z_2(z) e^{ik_z z}. \] (3.14)

Substituting 3.12 into 3.7, we obtain the final expression for \( \Omega_{pol,T} \):

\[ \Omega_{pol,T}^2 = \frac{A^2}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} dk_{\parallel} dk_z k_{\parallel} \frac{|F_{\parallel}^{(T)}(k_{\parallel})|^2 |F_{z}^{(T)}(k_z)|^2}{k_z^2 + k_{\parallel}^2}. \] (3.15)

### 3.2.3 Extension to two transitions

We next consider a system with two intersublevel transitions, both of which interact coherently with LO-phonons. These will later be identified as the nonradiative depopulation ("N", \( \psi_L \rightarrow \psi_I \)) and radiative lasing ("R", \( \psi_U \rightarrow \psi_L \)) transitions, respectively.

In considering more than one transition, it is found that we can define a particular mode associated with each. However, the problem arises that these modes are not generally the same nor orthogonal to one another. The transitions of our concern would couple to phonon modes \( N \) and \( R \), respectively, but we can choose instead basis modes \( N \) and \( \alpha \), where \( \alpha \) is a mode in the \( NR \)-plane of the mode space, but orthogonal to \( N \). This amounts to an orthonormalization within the basis of modes \( N \) and \( R \), which can be performed on a larger number of modes using a Gram-Schmidt process. A visualization of this is given in Fig. 3.3.

The matrix element for any electronic transition coupling to any arbitrary phonon mode \( Q \) is:

\[ \langle \psi_1; 0 | \hat{H}_f | \psi_2; 1_Q \rangle = \Omega_{pol,T} \langle T \cdot Q \rangle, \] (3.16)

where \( \langle T \cdot Q \rangle \) is the normalized inner product of mode \( Q \) with the mode associated with the transition. Therefore, the important matrix elements governing the problem when consider-
Figure 3.3: Orthonormalization within the mode space of two modes $N$ and $R$: the orthogonal modes are $N$ and $\alpha$.

Phonon modes $N$ and $\alpha$ are given for the $N$ transition as:

\[
\langle \psi_L; 0|\hat{H}_f|\psi_I; 1_N \rangle = \Omega_{\text{pol},N} \\
\langle \psi_L; 0|\hat{H}_f|\psi_I; 1_\alpha \rangle = 0,
\]

and for the $R$ transition as:

\[
\langle \psi_U; 0|\hat{H}_f|\psi_L; 1_N \rangle = \Omega_{\text{pol},R}\langle R \cdot N \rangle \\
\langle \psi_U; 0|\hat{H}_f|\psi_L; 1_\alpha \rangle = \Omega_{\text{pol},R}\langle R \cdot \alpha \rangle.
\]

The inner product $\langle R \cdot N \rangle$ can be computed by a sum over phonon modes as:

\[
\langle R \cdot N \rangle = \frac{1}{\Omega_{\text{pol},R}\Omega_{\text{pol},N}} \sum_{\vec{k}} F_{\vec{k},R} F_{\vec{k},N}^*,
\]

or as an integral by:

\[
\langle R \cdot N \rangle = \frac{A^2}{(2\pi)^3 \Omega_{\text{pol},R}\Omega_{\text{pol},N}} \int d^3\vec{k}\mathcal{F}^{(N)}(\vec{k})^* \mathcal{F}^{(R)}(\vec{k})/k^2.
\]

For simplicity, we choose an overall phase for $\alpha$ such that $\langle R, \alpha \rangle$ is positive real. Then,
since α lies in the plane defined by R and N, we have that:

\[ \langle R \cdot \alpha \rangle = \sqrt{1 - |\langle R \cdot N \rangle|^2}. \]  (3.21)

As was done for the polaron coupling strength, we also wish to simplify 3.20 by assumption of cylindrical symmetry. Moving the integral over \( \vec{k} \) to cylindrical coordinates, we have:

\[ \langle R \cdot N \rangle = \frac{A^2}{(2\pi)^3 \Omega_{\text{pol},R} \Omega_{\text{pol},N}} \int_{-\infty}^{+\infty} \int_{0}^{2\pi} \int_{0}^{2\pi} \, d\phi_{\vec{k}} dk_{\||} dk_{z} \frac{\mathcal{F}^{(N)*}(k_{z}, k_{||}, \phi_{\vec{k}}) \mathcal{F}^{(R)}(k_{z}, k_{||}, \phi_{\vec{k}})}{k_{z}^2 + k_{||}^2}. \]  (3.22)

In this case since the two transitions share a state, using the form factors from the previous section 3.6 leads to:

\[ \mathcal{F}^{(N)*}(\vec{k}) \mathcal{F}^{(R)}(\vec{k}) = i^{m_{U}+m_{L}-2m_{L}} e^{-i(m_{U}-m_{I})\phi_{\vec{k}}} \left[ \mathcal{F}^{(R)}(k_{z}) \mathcal{F}^{(R)}_{||}(k_{||}) \right] \left[ \mathcal{F}^{(N)*}_{z}(k_{z}) \mathcal{F}^{(N)*}_{||}(k_{||}) \right]. \]  (3.23)

We see the conservation of angular momentum rule again here: because of the integral over \( \phi_{\vec{k}} \), the quantity \( \langle R | N \rangle \) will vanish unless \( m_{U} = m_{I} \). This is stating that the phonon modes for two transitions which share a state only overlap if the other two states have the same angular momentum (which in this special case means that the two transitions have the same angular momentum, as stated before). The final expression for \( \langle R | N \rangle \) is then:

\[ \langle R | N \rangle = \delta_{m_{I},m_{U}} \frac{A^2}{(2\pi)^2 \Omega_{\text{pol},R} \Omega_{\text{pol},N}} \int_{-\infty}^{+\infty} \int_{0}^{2\pi} \, dk_{||} dk_{z} k_{||} k_{z}^2 \frac{\mathcal{F}^{(R)}(k_{z}) \mathcal{F}^{(R)}_{||}(k_{||}) \mathcal{F}^{(N)*}_{z}(k_{z}) \mathcal{F}^{(N)*}_{||}(k_{||})}{k_{z}^2 + k_{||}^2}. \]  (3.24)
3.2.4 Phonon decay and generation

LO-phonons have a finite lifetime due to anharmonic decay, typically into pairs of acoustic phonons [155]. A rigorous computation of the relaxation time $\tau_r$ for the LO-phonon towards equilibrium in spherical quantum dots was performed by Li and Arakawa [154]. It was found to be only weakly size-dependent for GaAs dots of diameters greater than 15 nm, and results were nearly identical for the two modes considered. In our model, we will use the approximate fit to their results for all modes:

$$\tau_r(T) = \left[ 8 - \frac{T}{54.5 \text{ K}} \right] \text{ps.} \tag{3.25}$$

$\tau_r$ involves both the competing decay and generation processes, where by detailed balance the two are equal and opposite at thermal equilibrium. Specifically, it is defined as

$$\frac{1}{\tau_r} \equiv -\frac{\Gamma^+ - \Gamma^-}{\delta N}, \tag{3.26}$$

where $\Gamma^\pm$ are the generation and decay rates and $\delta N$ is the deviation from equilibrium. We are interested in the bare decay and generation rates $\Gamma^\pm_n$, which are the transition rates between number states $n$ as depicted in Figure 3.4 ($\Gamma^\pm_n = 1/\tau^\pm_n$).

$\Gamma^+$ and $\Gamma^-$ come from Fermi’s Golden Rule:

$$\Gamma^\pm = \frac{2\pi}{\hbar} \sum_{\vec{k}, \vec{k}'} |H'_\pm|'^2 \delta(\hbar \omega - \hbar \omega' - \hbar \omega'') \tag{3.27}$$

Equation 3.27 includes a sum over all pairs of acoustic-mode wavevectors $(\vec{k}', \vec{k}'')$, where $H'_\pm$ is the pertubation Hamiltonian for each three-phonon interaction, in that way adding the transition rates involving all possible pairs. Direct evaluation of 3.27 is a complicated integral, which we will avoid dealing with directly. We do have, however, that:
Figure 3.4: LO-phonon decay and generation as transition rates between number (Fock) states of a mode.

\[
\Gamma^+ \propto N + 1 \\
\Gamma^- \propto N,
\]  

(3.28)

with \(N\) being the average occupation number. Importantly, the constants of proportionality are not the same. The actual rates depend also on the occupations of the acoustic phonons, but in a more complicated way. By detailed balance, the faster of the two rates depends on which side of equilibrium we are on, as is reflected in the definition of \(\tau\).

From 3.28 we can relate all \(\Gamma^+_n\) to each other, and separately all \(\Gamma^-_n\):

\[
\Gamma^-_n = n\Gamma^-_1 \\
\Gamma^+_n = (n + 1)\Gamma^+_0.
\]  

(3.29)

The bigger challenge is relating \(\Gamma^+_n\) to \(\Gamma^-_n\). Applying detailed balance to the ground state, we enforce that its population is not changing in time:
\[ \frac{dN_0}{dt} = -N_0 \Gamma_0^+ + N_1 \Gamma_1^- = 0 \] \tag{3.30}

Solving for the ratio \( \Gamma_0^+/\Gamma_1^- \), and then assuming a thermal equilibrium, we find that:

\[ \frac{\Gamma_0^+}{\Gamma_1^-} = \frac{N_1}{N_0} e^{-\beta E_{LO}} \] \tag{3.31}

This relationship will fix the relationships between all \( \Gamma_n^\pm \), but has only been checked to satisfy detailed balance for the ground state. For any other state \( n > 0 \):

\[ \frac{dN_n}{dt} = -\Gamma_n^- N_n - \Gamma_n^+ N_n + \Gamma_{n-1}^+ N_{n-1} + \Gamma_{n+1}^- N_{n+1} \] \tag{3.32}

\[ = -n \Gamma_1^- N_n - (n + 1) \Gamma_0^+ N_n + n \Gamma_0^+ N_{n-1} + (n + 1) \Gamma_1^- N_{n+1} \] \tag{3.33}

Next, we substitute in Equation 3.31, obtaining:

\[ \frac{dN_n}{dt} = \Gamma_1^- \left[ -n N_n - (n + 1)e^{-\beta E_{LO}} N_n + n e^{-\beta E_{LO}} N_{n-1} + (n + 1)N_{n+1} \right] \] \tag{3.34}

Finally, we can substitute in the assumption of thermodynamic equilibrium (\( N_{n-1} = e^{\beta E_{LO}} N_n \) and \( N_{n+1} = e^{-\beta E_{LO}} N_n \)), to find that:

\[ \frac{dN_n}{dt} = \Gamma_1^- N_n \left[ -n - (n + 1)e^{-\beta E_{LO}} + n + (n + 1)e^{-\beta E_{LO}} \right] \] \tag{3.35}

\[ = 0. \] \tag{3.36}

This result means that our set of relationships between rates given by Equations 3.29 and 3.31 is completely consistent with a thermodynamic equilibrium.

The next step is to compute all \( \Gamma_n^\pm \) absolutely, in terms of the relaxation rate towards
equilibrium \( (\tau) \). Starting from the definition of \( \tau_r \), we have the difference between the decay and generation rates:

\[
\Gamma_n^- - \Gamma_n^+ = (n - n_{LO}) \frac{1}{\tau} \quad (3.37)
\]

Substituting 3.29, we have:

\[
n \Gamma_n^- - (n + 1) \Gamma_0^+ = (n - n_{LO}) \frac{1}{\tau} \quad (3.38)
\]

Next substituting 3.31, we have:

\[
\Gamma_n^- [n - (n + 1)e^{-\beta E_{LO}}] = (n - n_{LO}) \frac{1}{\tau} \quad (3.39)
\]

Performing the algebra on the inside of the [], we obtain:

\[
n - (n + 1)e^{-\beta E_{LO}} = (1 - e^{-\beta E_{LO}})(n - n_{LO}). \quad (3.40)
\]

This factor of \( (n - n_{LO}) \) cancels with the one on the RHS of 3.39, yielding:

\[
\Gamma_n^- = \frac{1}{\tau} \frac{1}{1 - e^{-\beta E_{LO}}} \quad (3.41)
\]

From Equation 3.41, we can substitute the relationships between all \( \Gamma_n^\pm \), and we find that:

\[
\Gamma_n^- = \frac{1}{\tau_r} \frac{n}{1 - e^{-\beta E_{LO}}} \quad (3.42)
\]

\[
\Gamma_n^+ = \frac{1}{\tau_r} \frac{n + 1}{e^{\beta E_{LO}} - 1} = \frac{1}{\tau_r} (n + 1)n_{LO} \quad (3.43)
\]

Figure 3.4 shows the computed transition times as a function of temperature, for \( E_{LO} \) in GaAs (36 meV). At low temperature, the generation rates are extremely slow, and the relax-
ivation is dominated by decay, whereas at higher temperatures the generation does becomes important but remains significantly slower than the decay.

### 3.3 Application to a model system

#### 3.3.1 System and assumptions

We are now in the position to compute the steady-state transport and gain characteristics of a model QDCL. We choose perhaps the simplest possible system - a two dot module containing three electronic states. We treat the lateral quantum confinement as an infinite cylindrical potential, which allows for separation of variables between the axial and lateral dimensions. This could approximate for example the confinement of etched nanopillars, or nanowires growth from the bottom-up. In order to keep the problem tractable, we only consider the case where the lateral quantum confinement is sufficient so that only the lowest lateral energy state is relevant. In practice, this would require lateral confinement which is strong enough that the \( s - p \) energy separation is significantly above \( E_{LO} \). In GaAs, we obtain a value of 50 meV for this energy separation in a confinement diameter of 20 nm.

The bandstructure in the growth direction for this model system, in GaAs/Al\(_{0.2}\)Ga\(_{0.8}\)As,
adapted from [129, 156], is shown in Fig. 3.5. This design uses tunneling to inject electrons from the injector state $\Psi_I$ (black dashed) to the upper state $\Psi_U$ (green), after which there is a diagonal radiative transition between this state and the lower state $\Psi_L$ (red). The description “diagonal” refers to the fact that the states contributing to lasing are somewhat spatially separated, even though they still overlap. This is done intentionally for the sake of reducing the strength of the LO-phonon interaction across this transition. Electrons which do end up in the lower state are depopulated quickly by phonon emission downwards to the next injector state, so this transition is ordinarily placed close to $E_{LO}$ to make this as efficient as possible (known as “resonant-phonon” depopulation). As will be discussed in detail in the next Chapter, this strategy should be somewhat re-evaluated in the context of QDCLs.

Under the assumption of strong lateral confinement, the electron wavefunctions are the Bessel functions discussed in Sec. 2.1.3, for only $m = 0$, $n_r = 0$ (ground lateral states): $\psi(z, \rho, \theta) = \psi_z(z) J_0(k_\parallel \rho)$, where $\psi_z$ is the axial (growth-direction) wavefunction, $J_0$ is the zeroth-order Bessel function, and $k_\parallel$ is an in-plane wave vector that matches the pillar wall boundary condition at the first Bessel zero.

The layers thicknesses in angstroms starting from the injection barrier are 37/82/38/168, which at injection anticrossing gives a lasing transition of 10.6 meV (2.56 THz), a phonon depopulation transition of 36 meV, and a dipole matrix element for the optical transition of 4.7 nm. The injection tunnel coupling is found to be 1.9 meV, by taking one half of the injection anticrossing gap in a two-module bandstructure solution.

3.3.2 Tensor product Hilbert space and interactions

Given the three electronic states per module, a tensor product basis is built of the electronic state and the state of the two particular phonon modes that interact with the depopulation and radiative transitions ($N$ and $R$), orthonormalized by the procedure described in Sec. 3.2.3 to give orthogonal modes $N$ and $\alpha$. In Chapter 4, this will be further expanded to include a third phonon mode interacting with the $\psi_U \leftrightarrow \psi_I$ transition. We allow the total
number of phonons in both modes to reach up to two. At design biases, we can safely neglect coupling to the higher energy parasitic states $\psi_P$, and so it is not considered as part of our calculation.

The product states are coupled by the electron-phonon and the tunneling interactions, shown schematically in Figure 3.6. The red arrows represent coupling via $N$-phonons, and the green arrows coupling via $\alpha$-phonons. Blue arrows are the tunneling interaction, which only couples states having the same mode occupation. Not shown is a parasitic tunnel coupling, which will be neglected until Sec. 3.4.5. The vertical axis represents energy, although states grouped together are degenerate. Dashed arrows represent coupling between phonon numbers 1 and 2, which have a strength of $\sqrt{2}$ times those between 0 and 1. The module boundary is defined at the tunnel coupling, between the injector and upper electronic states.

It should be noted here that each module in reality contains its own pair of phonon modes $N$ and $\alpha$, and so by constructing our schematic of interactions as shown in Figure 4, we are implicitly enforcing that the occupations in all modules are perfectly correlated. This is of course not the case in a real system; however, this approximation is necessary in order to make the problem tractable. With faster dephasing, coherences spanning the entire module
are reduced, making the approximation closer to exact.

The values relevant to the electron-phonon interaction were computed as $\Omega_{\text{pol}, R} = 2.5 \text{ meV}$, $\Omega_{\text{pol}, N} = 3.3 \text{ meV}$, and $\langle R \cdot N \rangle = 0.176$.

### 3.3.3 Pure dephasing

A critical parameter is the pure dephasing time $T_2^*$, which encompasses all processes which decohere the various interactions without changing level populations. It contributes to the linewidth broadening for various transitions (for example, for a two level system the transition linewidth is increased by $2\hbar/T_2^*$), and also determines the coherence of the various interactions (for example, between states tunnel coupled by $\Omega$ if $\hbar/T_2 \gg \Omega$ the interaction will be incoherent, whereas if $\hbar/T_2 \ll \Omega$ it will be coherent, exhibiting strong coupling where the two states form an anticrossed doublet. $T_2^*$ adds to the total dephasing $T_2$).

Theoretical and experimental work suggests that decoherence in quantum dots occurs primarily due to both real and virtual acoustic phonon processes [157, 158]. In [157], $T_2^*$ was measured in InAs self-assembled quantum dots via four-wave mixing; values ranged from 90 ps at 10 K to 9 ps at 120 K. Dephasing was observed to be strongly temperature dependent (more so than the sublevel lifetimes), but also connected to the detailed energy structure of the system. As expected, however, these times are much longer than in conventional quantum-well QC-lasers, where $T_2^* \sim 300 \text{ fs}$. [134, 159] To simplify this intricate problem, we use a single phenomenological $T_2^*$ parameter throughout our simulations. Unless otherwise specified, this Chapter will assume $T_2^* = 5 \text{ ps}$ at 300 K, which is a reasonably conservative value and consistent with the computed values of [158].

### 3.4 Results

#### 3.4.1 Polaron-split gain

Results for the steady-state gain profile and population inversion at 100 K and 300 K are shown in Figure 3.7 for vanishing optical intensity. $\omega_0$ is the design frequency ($\hbar \omega_0 =$
Figure 3.7: Computed gain profiles under vanishing optical intensity at the injection anticrossing bias for different pure dephasing times, at 100 K and 300 K. Colored numbers on the left are inverted population fractions.

$E_U - E_L$). The electron density was taken to be $N_d = 10^{16} \text{ cm}^{-3}$, which corresponds to an active medium made up of a nanopillar array spaced on an 80 nm grid and doped with one electron per well. Pure dephasing times $T_2^*$ of 1 ps, 5 ps, and $\infty$ were applied to all coherences. $T_2^*$ is especially crucial for the peak gain and linewidth at low temperature, where the lifetimes of the $(n_N, n_\alpha) = (0, 0)$ states are extremely long due to the slow generation rate. At both temperatures, it is also noted that reduction in peak gain due to dephasing is attributed mainly to broadening rather than actual loss of population inversion.

Figure 3.8 shows gain profiles computed from each phonon occupation state separately at a temperature of 300 K and without pure dephasing for clarity. Even at 300 K, the large majority of the gain comes from the $(0, 0)$ states, which justifies our truncation at a total of two phonons. This is due to both their larger populations and longer lifetimes resulting in narrower linewidths. By separating the gain into occupation numbers we also note that while the total gain appears to exhibit five peaks, there are in fact more as well as resonance shifts which occur in the higher phonon occupation states.
3.4.2 Tunnel coupling dependence

The exact locations of the resonance peaks exhibit a complicated dependence on the coupling parameters and energy structure due to the complex nature of the chain-coupled problem, and also experience other shifts due to the damping mechanisms. However, these locations can be partially interpreted by diagonalizing $H$ only within the subspaces of states directly coupled to the zero-phonon radiative states, ignoring the phonon coupling across the radiative transition itself. The upper radiative states split into a doublet by the tunnel coupling back to the injector, described by the Hamiltonian $H_U$ in this subspace:

$$H_U = \begin{bmatrix} E_{rad} & \Omega_{tun} \\ \Omega_{tun} & E_{rad} \end{bmatrix} \rightarrow \text{eigenenergies } E_{rad} \pm \Omega_{tun}, \quad (3.44)$$

while the lower radiative state is split into a triplet by the phonon coupling to the next injector followed by the tunnel coupling to the next radiative state. The Hamiltonian $H_L$ in this three-level subspace is:
Figure 3.9: Zero-phonon gain at 300 K and \( T^*_2 = \infty \) as the tunnel coupling is turned on. Curves are offset by 200 cm\(^{-1}\) for clarity. Thin lines denote \( E_{\text{rad}} \pm \Omega_{\text{tun}} \) and \( E_{\text{rad}} \pm \Omega_{\text{tun}} \pm \sqrt{\Omega_{\text{tun}}^2 + \Omega_{\text{pol},N}^2} \).

\[
H_L = \begin{bmatrix}
0 & \Omega_{\text{pol},N} & 0 \\
\Omega_{\text{pol},N} & 0 & \Omega_{\text{tun}} \\
0 & \Omega_{\text{tun}} & 0
\end{bmatrix} \rightarrow \text{eigenenergies } 0, \pm \sqrt{\Omega_{\text{tun}}^2 + \Omega_{\text{pol},N}^2}.
\] (3.45)

\( E_{\text{rad}} \) is the energy of the radiative transition, where for the purpose of this section the energy of the lower radiative state is defined at zero. Given 3.44 and 3.45, the six possible transition energies are then \( E_{\text{rad}} \pm \Omega_{\text{tun}} \) and \( E_{\text{rad}} \pm \Omega_{\text{tun}} \pm \sqrt{\Omega_{\text{tun}}^2 + \Omega_{\text{pol},N}^2} \), all \( \pm \) combinations included.

Figure 3.9 shows the zero-phonon gain profile in comparison to these expected transition energies, as the injection coupling \( \Omega_{\text{tun}} \) is turned on. Adjusting \( \Omega_{\text{tun}} \) is equivalent to varying the thickness of the injection barrier. Simulations are performed at 300 K and with no pure dephasing. This model is sufficient at low tunnel coupling, where at \( \Omega_{\text{tun}} = 0 \) only the polaronic splitting exists, and the peak at the central frequency is absent due to its vanishing coupling strength to radiation. As the tunnel coupling strength is increased, the peak near center frequency begins to emerge and eventually dominates the gain profile. We attribute
the emergence of this peak to the onset of the phonon coupling across the radiative transition, which highly expands the Hilbert space relevant even to only the zero-phonon gain. This polaronic splitting represents a major difference compared to a conventional quantum-well QC laser, and must be properly account for in any design.

3.4.3 Phenomenological addition of a destructive scattering process

As the model up to this point included only electron tunneling, optical field, interaction with LO-phonons, decay of LO-phonons, and pure dephasing, one is led to wonder whether any additional effect might enter into the problem, altering the transport properties. Here, we consider a destructive electron scattering process across the radiative transition, which is modeled as a spontaneous boson emission rate $\tau_{sp}$, accompanied by stimulated emission and absorption rates $\tau_{st} = \tau_{abs} = \tau_{sp}/n_r$. $n_r$ is the Bose-Einstein occupation at the radiative energy. In an actual device, this might represent acoustic phonon scattering, for example. The gain remains relatively unaffected for $\tau_{sp} > 10$ ps, which is far faster than the expected speed of acoustic phonons or any other individual scatterer at these energies in QDs, indicating that the gain is quite robust to the presence of unexpected detrimental effects.

Figure 3.10: Gain at 300 K, $T_2^* = 5$ ps as a destructive scattering mechanism is introduced. Legend denotes values of $\tau_{sp}$. 
3.4.4 Gain saturation

A particular advantage to our method is the ability to automatically account for effects of increasing optical intensity directly onto the gain profile, allowing us to study gain saturation without needing to extract a stimulated emission rate. Figure 3.11 shows the change in gain profile as the circulating optical intensity is increased \( (I = 2\epsilon_0\nu c|E|^2) \), at 300 K and \( T_2^* = 5 \text{ ps} \). Reduction in peak gain is evident due to loss of overall population inversion, redistribution of population among various states, and effective lifetime broadening. In this way, the steady-state optical intensity could be estimated in a laser system by clamping the peak gain to the total cavity losses. Approximate saturation intensity is found to be on the order of 100-1000 W/mm², which amounts to 100 mW-1 W in a typical mode area of 10 \( \mu \text{m} \times 100 \mu \text{m} \). This would then be the approximate total intracavity intensity in both directions.

3.4.5 Subthreshold parasitic current channel

To this point, we have focused on somewhat of an ideality, where only tunnel coupling from the injector to upper radiative state is considered. However, it is well known that a major issue for THz QC-lasers is the existence of a parasitic current channel that occurs for voltage...
Figure 3.12: Effects of a parasitic tunneling channel at $T = 300$ K. Left: Transport characteristic for $T_2^* = 5$ ps and 1 ps. Right: Gain at various bias points for $T_2^* = 5$ ps. (labeled in mV/module and offset by 200 cm$^{-1}$ for clarity)

biases below the injection resonance [134, 58, 160]. While the details vary between designs, this current channel is associated with tunneling from the injector to the lower radiative state or the excited state in the wide depopulation well. The presence of this parasitic current sets a floor on the threshold current density, and if it is too strong, creates a premature NDR, which prevents reaching the design bias. In conventional QC-lasers, since this coupling is typically $\Omega_p \sim 0.2$-0.5 meV, the relatively fast dephasing ($T_2^* \sim 0.3$ ps) helps to suppress this current. Since the dephasing times in a quantum dot QC-laser are expected to be 1-2 orders of magnitude longer, a concern naturally arises that this parasitic channel will be too strong.

We now introduce the parasitic tunnel coupling from the injector to the lower radiative state, which has a value computed from the level anticrossing of $\Omega_p = 0.875$ meV. Although this channel is well detuned at the injection resonance, it is however important at lower bias. Figure 3.12 demonstrates the effect of the parasitic coupling on the transport characteristic for $T_2^* = 5$ ps and 1 ps, and the gain at various bias points for $T_2^* = 5$ ps. Very large current flow is found at biases over a wide range around the parasitic resonance, leading to a considerable NDR. As expected, the gain is significantly modified at lower bias points while
at higher bias the parasitic tunneling becomes unimportant as it is further detuned.

A similar current instability was predicted in [153], where the possibility of doubling all barrier thicknesses was explored. In our two-well design, where the radiative transition is diagonal, it is clearly disadvantageous to increase the radiative barrier, but for example doubling only the injector barrier thickness from 3.7 to 7.4 nm reduces the injection coupling from 1.9 to 0.3 meV and the parasitic coupling from 0.875 to 0.2 meV. However, this reduction in the injection coupling introduces other complications, importantly a large splitting in the gain spectrum as shown in Figure 3.9. Furthermore, in order to appreciably reduce the parasitic current level, one requires the coupling to be $\Omega_p \ll \hbar/T^*_2$ (0.13 meV for $T^*_2 = 5$ ps), which is difficult to achieve in this simple two-well design.

3.5 Analysis

This Chapter has analyzed the feasibility of an idealized QDCL naively adapted from an ordinary QCL design. Importantly, many realistic concerns such as quantum dot size inhomogeneity have not been accounted for, making for a prediction of specifically the effects of polaron formation on transport. A further limitation to the model has also been imposed by computational considerations; only two phonon-coupled transitions were accounted for, restricting our treatment to the regime of small pillar diameter ($\sim$20 nm in GaAs), where only the lateral $s$-states are important to the problem. The results in this Chapter therefore serve to elucidate the expected behavior of devices in which essentially perfect material control has been achieved, or in other words in the upper limit of anticipated device performance for the considered design.

A key motivating result is the prediction of easily sufficient gain for lasing at room temperature, on the order of 100 cm$^{-1}$ even for a fast estimate of pure dephasing (the losses in a low-loss metal-metal waveguide are roughly 15-30 cm$^{-1}$). However, the gain profile can be highly split to the level of several meV by the strong electron-LO-phonon coupling, and the spectral position of the gain peaks can be difficult to predict for significant levels of the tunnel coupling. This makes design strategies significantly more complicated to devise than
in well-based QCLs. However, still none of these effects are actually prohibitive of lasing action.

The effect that would prevent actual operation of the QDCL modeled in this Chapter is the presence of extreme subthreshold parasitic current. The subthreshold parasitic causes a negative differential resistance (NDR) in each module’s transport characteristic before the intended operating bias, which means that the operating condition calls for the biasing of hundreds of NDRs in series. This is highly electrically unstable; in reality electric field domains will form, with some modules biased on the initial positive different resistance (PDR), and others on a PDR of higher voltage, which is the stable operating condition in which current is continuous through the structure. As the bias is increased, new fields are not reached, but rather only transfer of additional modules to the higher-current PDR to accommodate the higher total bias. This effect is well-known to occur in quantum well superlattices [161] and occurs also in THz QCLs where there exists an NDR beyond the intended operating bias point [162]. The unfortunate expected effect in a QDCL is that the intended operating bias itself lies within the range of electrical instability.

The reason that such a strong current maximum occurs at the parasitic resonance is because transport through the structure at that point is simply repeated resonant tunneling followed by resonant phonon emission. This same current peak is sometimes visible in ordinary QCLs, but does not exceed the current at operating bias, where depopulation is made extremely fast by resonant LO-phonon emission and the optical transition does not present too much of a bottleneck to transport because of the fairly short upper state lifetime. In addition, the longer dephasing time in a QDCL allows for the long-range parasitic tunneling to emerge much more strongly than is observed in QCLs.

In summary, it has been found that, in the case of the QDCL, it is in fact the phonon bottleneck effect itself that serves to greatly suppress transport at the intended operating point, where there exists no state spaced closed to \( E_{LO} \) below the upper state, causing an electrical instability. Avoidance of this is a major challenge, as it is a result stemming from the very same effect that is hoped to enable gain at room temperature in the first place. The next Chapter will present a new and counterintuitive design strategy devised with this
in mind, and also account for the effect of size inhomogeneity in the quantum dots.
CHAPTER 4

Realistic Design Strategy For a Quantum Dot Cascade Laser

4.1 Introduction

The previous Chapter concluded that although THz QDCLs can in principle provide THz gain at room temperature, this gain will likely exist in an electrically unstable operating point, at least if conventional QCL designs are scaled to the quantum dot limit. This is due to an extreme subthreshold parasitic current channel that is efficient when the injector sublevel $I$ from the previous module aligns with the lower radiative state $L$. This effect would create a negative differential resistance (NDR) bias regime beyond the parasitic point. The NDR regime leads to space charge buildup and high-field domain formation, which effectively renders the intended operating bias point inaccessible, and prevents laser operation.

NDR occurs because transport is far more efficient at the parasitic bias compared to the higher design bias. This is a natural consequence of the lower radiative state having a short lifetime due to relaxation by LO-phonon emission; this occurs by design by setting the lower subband energy such that $E_{LI} \approx E_{LO}$. In conventional QCLs, this parasitic channel is kept modest by using thick injection barriers to ensure that the tunneling process — and not the lifetime $\tau_L$ — is the bottleneck to current flow [114]. In the QDCL, this strategy will not work, because the longer lifetimes and the longer dephasing times mean that barriers would have to be impractically large to suppress the parasitic channel.

In this Chapter, we present a counterintuitive strategy for the design of a THz QDCL which avoids the expected problem of electrical instability. This is in contrast to previous
theoretical investigations which have for the most part focused upon designs which follow the same design rules as conventional QCLs. The central concept is to engineer the sublevel spacings so that resonant-phonon emission occurs from the laser’s upper state, i.e. $E_{u1} \approx E_{LO}$, rather than from the lower state as is widely adopted in ordinary QCLs. This has the effect of ensuring the maximum current occurs at the design bias, rather than at below. The strategy is tested theoretically using the density matrix model presented in the previous Chapter extended to a level of generality so that both the original and new designs are tested using the same code. It is predicted that sufficient gain is available at room temperature and at electrically stable operating points, although the gain level and saturation intensity are lower than for the conventional strategy. Finally, simulations of quantum dot size inhomogeneity are performed, which suggest that some amount of size fluctuation does not necessarily prohibit lasing.

The Chapter is organized as follows: Section 2 gives an overview of the design concept including important points on the physics of electron-LO-phonon interaction in quantum dots, Section 3 overviews the density matrix model used to assess the new strategy, Section 4 gives the calculated results, Section 5 addresses the effect of quantum dot size inhomogeneity, and concluding remarks are made in Section 6.

4.2 Design concept

4.2.1 Key polaron physics

As explained in detail in the previous Chapter, the discrete nature of electronic sublevels in quantum dots combined with the nearly dispersionless LO-phonon spectrum results in a strong coupling of the two and the formation of electron-phonon quasiparticles known as “polarons”. Because the combined system does not have a continuum of states, irreversible decay does not occur. Instead, the electron continually emits and re-absorbs an LO-phonon, resulting in a Rabi oscillation which continues until the decay of the LO-phonon or another interruption [87, 151]. Some reiteration of the physics will be given here to aid the explanation
It can be shown that a given intersublevel electronic transition emits phonons into only a single mode spatially localized around the transition. The polaron coupling strength to this mode is defined as $\Omega_{\text{pol}}$, which couples for example the product states $|\psi_2; 0\rangle$ and $|\psi_1; 1\rangle$, which are formed from the electron states $\psi_n$ and phonon mode occupations $m$. This is illustrated qualitatively in Fig. 4.1(a), where we consider the energy eigenstates obtained by diagonalization of the electron-phonon Hamiltonian in a fictitious InAs quantum dot of variable dimension such that the intersublevel energy spacing $E_2 - E_1$ is varied. The dashed lines are the state energies without the electron-phonon interaction; inclusion of the interaction causes an anticrossing behavior between the polaron eigenstates. Full hybridization occurs for an electronic transition tuned to $E_{\text{LO}}$ (29 meV in InAs), whereas for detuned transitions the eigenstates return more so to their separate “bare” characters, i.e. “electron-like” or “phonon-like.”

LO-phonons can irreversibly decay into various pairs of acoustic phonons, as long as the total momentum and energy of the pair is conserved. The overall relaxation rate can be computed by integrating over the continuum of possible pairs, and has been performed in various works [151, 154]; here we borrow the results from [154] and approximate a 2.5 ps
relaxation time to equilibrium at room temperature. From this rate, the transition rates between the phonon number (Fock) states of a given mode, which are the decay and generation of phonons, are deduced.

Fig. 4.1(b) shows the approximate relaxation dynamics for three different intersublevel energies: $E_{21} = 10$, 20, and 29 meV. Two key characteristics are observed. First, despite the existence of Rabi oscillations, the envelope of the relaxation process can be described by an approximate lifetime. Second, this lifetime increases for smaller $E_{21}$, as the intersublevel transition becomes further detuned from $E_{LO}$, and the polaron becomes less "phonon-like". Fig. 4.1(c) repeats the example of Fig. 4.1(b), only now including a finite pure dephasing time $T_2^*$. $T_2^*$ in quantum dots has been investigated experimentally using a four-wave mixing experiment in [157], where it was found to decrease from approximately 90 ps at 10 K to 9 ps at 120 K. Calculations in [158] indicate that $T_2^*$ is intricately dependent on the details of the energy structure, but expected values are a few ps at 300 K. In this work we will use $T_2^* = 3$ ps, phenomenologically, which is an order of magnitude longer than typical values assumed for $T_2^*$ in quantum wells. As seen in Fig. 4.1, the presence of pure dephasing has a fairly small effect on transport, although the effect on the optical linewidth is more significant; further discussion is found in the previous Chapter.

4.2.2 Design strategy

The most successful THz QCL designs are based on the “resonant-phonon” concept, which is shown in Fig. 4.2(a) for a simple 3-level laser design (adapted from [129, 132]), with upper radiative, lower radiative, and injector states labeled $U$, $L$, and $I$ respectively. Fast depopulation is ensured by designing $E_L - E_I \approx E_{LO}$.

We propose the new concept specifically for QDCL design where it is the upper, rather than the lower, state that is placed to depopulate at resonance with LO-phonons. In other words, the design is such that $E_U - E_I \approx E_{LO}$. This is shown in the design of Fig. 4.2(b), which is adapted from the previous design simply by widening the wells, in contrast to the suggestion in [153] of widening the barriers. This would certainly not be a useful concept
Figure 4.2: Electronic bandstructures and level schematics at parasitic/design biases for (a) a conventional QCL design and (b) a modified design for a QDCL. The chosen material system is InAs/InAsP and the layer thickness in nm starting with the tunnel barrier are 5.5/14/3/30 for (a) and 5.5/18/3/35 for (b), so that the only difference is a widening of the wells. The level schematics illustrate the design strategy which is to engineer resonant LO-phonon depopulation of the upper state $\Psi_U$, rather than the lower state $\Psi_L$, to the injector state $\Psi_I$. 

(a) Conventional

(b) Modified
in ordinary QCL design, as the upper state lifetime would most likely be far too short to support sufficient population inversion at a reasonable current density. However, here it is imperative that the current maximum occurs at the intended design bias (where \( I' \) injects into \( U \)) rather than at a lower bias (where \( I' \) injects into \( L \)). And indeed, since \( E_L - E_I < E_{LO} \), depopulation is slowed. However, all of these things notwithstanding, it will be shown that the laser can in fact still support a sufficient population even under the condition that the effective lifetime of the upper state is shorter than the lower.

The condition for electrical stability is that the current at design bias \( J_0 \) should be greater than the current at the parasitic resonance \( J_{para} \). Although a real estimation of the current at either point requires a detailed simulation, the design strategy that we employ can be understood by thinking of the polaron relaxations as leading to effective relaxation lifetimes \( \tau_U \) and \( \tau_L \) (similar to shown in Fig. 4.1). Since the primary bottleneck to transport in a QDCL is the polaron relaxation, the injector level can be expected to share population with the upper/lower state when biased to resonance with either, if the further approximation of perfectly selective injection is made. This leads to an estimation of the currents as:

\[
J_{para} \approx \frac{N_s q}{2\tau_L} \quad J_0 \approx \frac{N_s q}{2\tau_U},
\]

(4.1)

where \( N_s \) is the sheet doping density. It is understood through Eq. 4.1 that the condition \( \tau_U < \tau_L \) is required for electrical stability.

Furthermore, the concept of rates would give the steady-state population inversion in our system as:

\[
\Delta N = \frac{J}{q} \frac{1 - \frac{\tau_L}{\tau_{UL}}}{\tau_U},
\]

(4.2)

which establishes the simple condition that the lower state lifetime \( \tau_L \) need only be shorter than the transition time across the radiative states \( \tau_{UL} \) to yield positive population inversion. A short upper-state lifetime \( \tau_U \) reduces the magnitude of the population inversion, but importantly does not change the sign. We expect that the condition \( \tau_{UL} > \tau_L \) can be
satisfied in our system as long as the depopulation energy $E_L - E_I$ is tuned closer to $E_{LO}$ than the radiative energy $E_U - E_L$.

### 4.3 Extension of the model to three phonon modes

As was the case in the previous Chapter, we are again assuming that the pillar diameter is sufficiently small that only the lateral ground states are relevant to the transport problem. The pillar cross-section is approximated as circular with a diameter of 30 nm, which results in a separation of 66 meV between the ground and first excited lateral states in InAs, well above the LO-phonon energy of 29 meV. The phonon interaction is now described using three orthogonalized phonon modes which span the space defined by the three possible intramodule transitions: $U-L$, $U-I$, and $L-I$. This is an extension of the simulation in the previous Chapter, which only included two modes ($U-L$ and $L-I$), and allows for both designs to be simulated using the exact same code. The maximum total occupation in all modes is again allowed to reach up to two — sufficient to capture the relevant phenomena at room temperature. A schematic of the full electron-phonon product space with electron-phonon interactions drawn

---

**Figure 4.3**: Schematic of electron-phonon interactions in the tensor product Hilbert space when three phonon modes are considered. Orange arrows labeled by Roman numerals represent groups of couplings.
is shown in Fig. 4.3. This is an extension Fig. 3.6 to include three phonon modes instead of only two.

The necessary parameters needed for the mode orthonormalization are the nine inner products between all basis modes and all orthonormalized modes. The basis modes, defined by transitions $U-I$, $L-I$, and $U-L$, are named $x$, $y$, and $z$, respectively, and the orthonormalized set spanning this space is called $a$, $b$, and $c$. Performing a Gram-Schmidt orthonormalization procedure, the inner products we need can be expressed in terms of the inner products within the original set (which can be evaluated using 3.24):

\[
\begin{align*}
    a \cdot x &= 1 & a \cdot y &= 0 & a \cdot z &= 0 \\
    b \cdot x &= a \cdot b & b \cdot y &= \left[ 1 - (a \cdot b)^2 \right]^{1/2} & b \cdot z &= 0 \\
    c \cdot x &= a \cdot c & c \cdot y &= \frac{b \cdot c - (a \cdot b)(a \cdot c)}{b \cdot y} & c \cdot z &= \left[ 1 - (c \cdot x)^2 - (c \cdot y)^2 \right]^{1/2}.
\end{align*}
\]

The electron-phonon interaction can then be included into the simulation by using the following couplings, organized by groups defined in Fig. 4.3. The first and second column show the phonon part of the product states involved, and the third column is the coupling strength in terms of $\Omega_x$, $\Omega_y$, $\Omega_z$, and the inner products found using 4.3 ($\Omega_n$ is the polaron coupling strength of transition $n$):
<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>000 100 $\Omega_x a \cdot x$</td>
<td>000 100 $\Omega_x b \cdot x$</td>
<td>000 100 $\Omega_x c \cdot x$</td>
</tr>
<tr>
<td></td>
<td>000 010 $\Omega_y a \cdot y$</td>
<td>000 010 $\Omega_y b \cdot y$</td>
<td>000 010 $\Omega_y c \cdot y$</td>
</tr>
<tr>
<td></td>
<td>000 001 $\Omega_z a \cdot z$</td>
<td>000 001 $\Omega_z b \cdot z$</td>
<td>000 001 $\Omega_z c \cdot z$</td>
</tr>
<tr>
<td>IV</td>
<td>100 200 $\sqrt{2}\Omega_x a \cdot x$</td>
<td>100 200 $\sqrt{2}\Omega_x b \cdot x$</td>
<td>100 200 $\sqrt{2}\Omega_x c \cdot x$</td>
</tr>
<tr>
<td></td>
<td>100 110 $\Omega_y a \cdot y$</td>
<td>100 110 $\Omega_y b \cdot y$</td>
<td>100 110 $\Omega_y c \cdot y$</td>
</tr>
<tr>
<td></td>
<td>100 101 $\Omega_z a \cdot z$</td>
<td>100 101 $\Omega_z b \cdot z$</td>
<td>100 101 $\Omega_z c \cdot z$</td>
</tr>
<tr>
<td></td>
<td>010 110 $\Omega_x a \cdot x$</td>
<td>010 110 $\Omega_x b \cdot x$</td>
<td>010 110 $\Omega_x c \cdot x$</td>
</tr>
<tr>
<td></td>
<td>010 020 $\sqrt{2}\Omega_y a \cdot y$</td>
<td>010 020 $\sqrt{2}\Omega_y b \cdot y$</td>
<td>010 020 $\sqrt{2}\Omega_y c \cdot y$</td>
</tr>
<tr>
<td></td>
<td>010 011 $\Omega_z a \cdot z$</td>
<td>010 011 $\Omega_z b \cdot z$</td>
<td>010 011 $\Omega_z c \cdot z$</td>
</tr>
<tr>
<td></td>
<td>001 101 $\Omega_x a \cdot x$</td>
<td>001 101 $\Omega_x b \cdot x$</td>
<td>001 101 $\Omega_x c \cdot x$</td>
</tr>
<tr>
<td></td>
<td>001 011 $\Omega_y a \cdot y$</td>
<td>001 011 $\Omega_y b \cdot y$</td>
<td>001 011 $\Omega_y c \cdot y$</td>
</tr>
<tr>
<td></td>
<td>001 002 $\sqrt{2}\Omega_z a \cdot z$</td>
<td>001 002 $\sqrt{2}\Omega_z b \cdot z$</td>
<td>001 002 $\sqrt{2}\Omega_z c \cdot z$</td>
</tr>
</tbody>
</table>

### 4.4 Results

Simulated results of the transport characteristics and bias-dependent small-signal optical gain at 300 K are in Fig. 4.4. It is assumed for both designs that the pillars are spaced on a square lattice of 60 nm pitch, and that one electron exists per module, amounting to an overall electron density of approximately $5 \times 10^{15}$ cm$^{-3}$. Due to the very large CPU time required, these simulations make use of highly-parallel (high-throughput) computing.

The strong parasitic channel is evident in the conventional design at $\sim$25-30 mV/module, with a current density approximately 5× larger than at the design bias of 42 mV/module. In contrast, while the modified design does exhibit a parasitic current channel at 25 mV/module, the current is much larger at the design point of 29 mV/module. The conventional design is calculated to exhibit higher gain at design bias than the modified design, but the bias point is unstable making this gain entirely inaccessible. In fact, hardly any gain at all is predicted in the conventional design at any stable points. The modified design, on the other
Figure 4.4: Simulated transport characteristics and bias-dependent small-signal optical gain at 300 K for the conventional and modified designs, showing regions of electrical stability/instability. The designed and predicted gain in the modified design is circled by the dashed oval.
hand, still produces a significant amount of gain (on the order of 100 cm$^{-1}$) in the region of stability and around the frequency that is visible from the electronic bandstructure. This result suggests that our concept is valid.

Effects of gain saturation are investigated in Fig. 4.5. In these simulations, the small-signal gain at frequency $\omega$ is calculated under increased intensity at the frequency of peak small-signal gain $\omega_{\text{peak}}$, so that cross-saturation of $\omega$ by intensity at $\omega_{\text{peak}}$ is obtained. Extension of the model to include multiple frequencies was performed in manner similar to [133]. Results suggest that in both designs, the saturation is highly nontrivial owing to the complex nature of the polaron-coupled system; aspects of both homogeneous and inhomegeneous saturation are visible. Importantly, spectral hole burning effects are important even for perfectly uniform quantum dot size. Fig. 4.5(c) shows the small-signal gain spectrum peak as a function of saturation by intensity at $\omega_{\text{peak}}$. The modified design exhibits a saturation
intensity around the 10 W/mm² level, which is approximately one order of magnitude lower than the conventional design. As was done for the current (4.1) and gain (4.2), this can also be understood in terms of effective rates; the saturation intensity can be written for a three-level laser as:

$$I_{sat}(\nu) = \frac{h\nu}{\sigma(\nu) \left[ \tau_L + \tau_U \left( 1 - \frac{\tau_L}{\tau_U} \right) \right]}.$$  \hspace{1cm} (4.4)

where $\nu$ is the optical frequency and $\sigma(\nu)$ is the gain cross-section. The low saturation intensity is caused by the long lower state lifetime, which is a consequence of $E_L - E_I < E_{LO}$.

### 4.5 Effects of inhomogeneity

Because quantum dot sizes cannot be controlled at the atomic precision level as is possible in quantum wells, size inhomogeneity can be expected to be an important factor in device operation, quite possibly dominating the optical linewidth. In the case of self-assembled quantum dots the growth process is inherently stochastic. Quantum dots defined by bottom up nanopillar growth and top-down etching have lateral dimensions defined by lithography — nonetheless some nonuniformity will likely persist. In all cases, due to the complexity of the gain profile which is split by the nontrivial interplay of tunneling and polaronic couplings, assessment of the inhomogeneous broadening cannot be simply assumed to follow a normal distribution.

To approximate the effect of inhomogeneity, simulations were run in which all length dimensions (4 layer widths and the diameter) were allowed to fluctuate by a Gaussian distribution of 5% standard deviation. The small-signal gain is computed for 50 variations on each design, which are all assumed to be biased at the design bias of the target geometry. This would ignore the effects of, for example, spatial correlation in dimensions that might be expected in the growth reactor, and the slightly different biases that each might see due to series combination. However, this still provides a rough look at the sensitivity of the gain profile to fluctuations in quantum dot dimensions. Results are shown in Fig. 4.6. The 5%
standard deviation in lengths is expected to reduce the peak gain by approximately one half, although this would remain well above a typical cavity loss of 25 cm$^{-1}$.

### 4.6 Conclusion

A useful strategy was presented for the design of a THz QDCL that avoids the predicted problem of electrical instability. The concept is, counterintuitively, to aim for resonant-phonon depopulation of the laser’s upper, as opposed to lower, state. This creates a parallel current channel at design bias from $U \rightarrow I$ that ensures one can reach design bias. This can be considered conceptually equivalent to the strategy of using a shunt resistance to allow bias of a superlattice in the NDR regime so as to obtain Bloch gain [163]. Although somewhat lower gain is predicted in the modified design as compared to a conventional design hypothetically at design bias, this is still at the level of greater than 100 cm$^{-1}$ which should be easily sufficient for lasing. A downside brought about by our modification is an order-of-magnitude lower saturation intensity because of the long lower state lifetime, which could lead to a lower power level. Finally, simulations of quantum dot size inhomogeneity were performed, which suggests that at least a 5% uncertainty in the dot dimensions is tolerable. We should also note that this strategy will likely work only when $E_U - E_L < E_L - E_I$, such
that the radiative transition is further from $E_{LO}$ than the depopulation. In InAs dots for example, this will limit the photon energy to at most 14 meV ($\sim$3.5 THz).

These results provide a concrete example of the contrast between successful QCL and QDCL design concepts, which is called for because of the inherently different transport physics. Further work could focus on more complicated designs involving a larger module, although this would be outside the practical capability of this modeling approach, or on the inclusion of charging effects [164].
CHAPTER 5

Terahertz Difference Frequency Susceptibility in Midinfrared Quantum Cascade Lasers

Although THz lasing has never been observed in a QCL at room temperature, coherent THz radiation has been demonstrated using nonlinear, rather than direct, generation. Mid-IR QCL active regions have been stacked in series combination inside the same laser cavity, where they are intended to lase simultaneously at separate frequencies. The difference frequency is in the THz, which can be produced by the mixing of the mid-IR frequencies coming from the optical nonlinearity of the active region. Because the mid-IR QCL operates at room temperature, coherent THz radiation at room temperature is produced.

This Chapter aims to provide a fresh look at the nature of the difference-frequency susceptibility in quantum wells, provided by our new density matrix transport formalism. Section 5.1 gives an overview on optical nonlinearity in quantum wells and a preview on the new insights, Section 5.2 outlines the extension of our modeling concepts for optical nonlinearity and provides examples of its capability, Section 5.3 shows how the scattering and dephasing rates are calculated, Section 5.4 gives the results on an example device, and Section 5.5 provides new analysis including paths forward to exploit the new findings. More information on the steady-state solution under nonlinear frequency-mixing is found in Appendix A.2.

Figure 5.1 shows an example of the two active regions in one such device from [44], which will be used as an example system throughout this Chapter.
Figure 5.1: Bandstructures in an example mid-IR QCL designed to emit THz radiation by difference frequency generation. The highlighted states participate the most in lasing action.

## 5.1 Background

### 5.1.1 Resonant susceptibility

It has been known that quantum wells possess very large optical nonlinearity since before the invention of the quantum cascade laser. This includes optical rectification [90, 165], harmonic generation [166, 167, 168], difference-frequency generation [169, 93], and four-wave mixing [170, 171]. Ordinarily, an estimation of the nonlinearity involves the use of “sum-over-states” (SOS) expressions derived from perturbation theory, an approach which is inherited from atomic physics to essentially perform a sum over contributions from all possible combinations of states, regardless of resonance [172]. For example, the SOS expression for difference-frequency generation is:
\[ \chi^{(2)}(\omega_1 = \omega_3 - \omega_2) = \frac{N_d q^3}{\hbar^2 \epsilon_0} \sum_{mn} z_{ln} z_{nm} z_{ml} (\rho_{ll}^{(0)} - \rho_{mm}^{(0)}) \]
\[ \times \left( \frac{1}{\omega_{nl} - \omega_1 - i\Gamma_{nl}} + \frac{1}{\omega_{nm} + \omega_1 + i\Gamma_{nm}} \right) \times \left( \frac{1}{\omega_{ml} + \omega_2 - i\Gamma_{ml}} + \frac{1}{\omega_{ml} - \omega_3 - i\Gamma_{ml}} \right). \]

(5.1)

This expression is a sum over all triplets of states \( l, m, n \), where \( z_{ab} \) are the dipole matrix elements, \( \rho_{aa}^{(0)} \) are the state fractional populations to zeroth order in optical field, \( \omega_{ab} \) are the resonant frequencies of transitions, and \( \Gamma_{ab} \) are the transition broadenings. Importantly, any terms satisfying \( n = l, n = m, \) or \( m = l \) are not intended to be included, which would call for the state centers \( z_{aa} \), as opposed to only dipoles \( z_{ab} \). In fact, it can be shown that the expression can be made translationally variant by including these terms. In the case of an atom, where all states are centered in the same location at the nucleus, this is not problematic because simply defining this location as zero makes these terms vanish, but in a QCL there is an important contrast because the differences in state centers \( z_{bb} - z_{aa} \) are quite often at the same level as the dipole matrix elements \( z_{ab} \). Implications of this, and other effects not covered in the SOS expression, are discussed in the next section.

The strategy that has been used so far to optimize the difference frequency susceptibility is to engineer a resonance between three states. This is similar to the engineering of a gain transition in a QCL, but extended to include three levels where the three possible transitions resonate with the optical frequencies involved in the nonlinear interaction. In fact, the resonant susceptibility is usually designed involving the gain transition(s) themselves, which gives the added benefit that the mid-IR pumps experience gain rather than absorption. This so-called “resonant susceptibility” could occur between, for example, states 10, 11, and 12 in Active Region B (see Fig. 5.2), or sometimes more than two triplets of states are expected to contribute (see highlighted states in Active Region A).

For the triplet shown in Fig. 5.2, picking out only the resonant terms in the SOS expression simplifies the susceptibility to \( \Delta N \approx N_{12} - N_{11} \approx N_{12} - N_{10} \) is the approximate
Figure 5.2: Depiction of the resonant difference frequency susceptibility present in Active Region B (Fig. 5.1), occurring across the inverted optical transitions.

It is convenient that optimization of the gain through maximizing population inversion and dipole matrix elements also helps to maximize the resonant susceptibility.

5.1.2 Previous theoretical work

The most simple model for the difference frequency susceptibility, which has been the most widely used to date, is to assume that the susceptibility is completely described by the resonant terms (Eq. 5.2), with the population inversion estimated by clamping the estimated gain to the cavity losses. The broadening parameters are usually estimated. This sort of simple estimation has led to predictions of $\chi^{(2)}$ at the level of 10s of nm/V, ever since the very first demonstration [93].

Evaluating the full SOS expression (5.1) calls for a more sophisticated transport model, because all of the state populations and transition broadenings must be known. For this, Ensemble Monte Carlo simulations have been performed which can calculate the whole
steady-state of the laser under lasing conditions, and resolved in-plane, using rate equa-
tion approximations [173, 174]. The full SOS expression is then applied given the knowledge of populations and broadening parameters coming from the detailed solution. These sim-
ulations have for the most part lent agreement to the notion that $\chi^{(2)}$ is dominated by the resonant processes around the lasing transition, at least for biases near the injection resonance.

5.2 Extension of density matrix concepts for nonlinear optics

The density matrix model exhibited throughout this Thesis presents an opportunity to study nonlinear optics in quantum wells in a more general way than has been done previously. At its base, its concept is simply to solve for the electron steady-state density matrix under harmonic excitation, which can be generalized to include multi-harmonic excitation (optical fields at more than one frequencies) and response at the resultant mixing frequencies. Modeling the electron motion in this way has the benefit that no *a priori* assumption is made as to the nature of the nonlinear interaction, allowing for the true character to emerge on its own. In addition, the modeling allows for a finite optical field to enter into the problem nonperturbatively, so that the electron motion at dc and ac included frequencies is solved all together in a fully-coupled sense. This Section overviews the mathematics required to extend the method for nonlinear optics, and provides examples of its generality by showing the range of effects that can emerge naturally.

5.2.1 Newly identified effects for THz difference frequency generation

Figure 5.3 illustrates three separate effects that this Chapter will demonstrate as important to THz difference frequency generation. The atom-like processes are the nonlinearity described by the SOS expression, which are strongest at the resonant condition drawn. This is akin to optical polarization of an atom. Permanent dipole effects involve systems where the state centers are different: an ac field at *any* frequency $\omega_3$ would be able to modulate the transition energy separation by a first-order Stark effect, which would then modulate the first-order
susceptibility $\chi^{(1)}$ seen by another ac field at $\omega_2$ near resonance. The end result is a second-order susceptibility $\chi^{(2)}$. The third effect that we identify will be called “self-detection,” but could also be called simply “photoconductivity”: this effect comes from an increase in the QCL drive current which follows the beatnote created by the superposition of the two mid-IR pump frequencies. This effect is well-known to occur in QCLs under only a single frequency excitation, where the dc current increases significantly with the lasing optical field; here we will demonstrate the the bandwidth of this effect reaches into the THz regime because of the subpicosecond intersubband scattering times.

Neither the permanent dipole nor self-detection effects are describable using the SOS expression. The permanent dipole effect would necessarily require the inclusion of diagonals of the $z$ operator, which leads to unphysical translational variance as will be shown in Section 5.2.4. The self-detection effect requires the treatment of the QCL as an infinite periodic system comprised of modules with conduction currents flowing between, which is not amenable to the SOS treatment which assumes an atom, which is only a finite-sized polarizable entity. Essentially, by making no a priori distinction between conduction and displacement currents, our model will show that both can in fact contribute to nonlinearity. At high frequency, this self-detection current should perhaps be considered in a “gray area” between conduction and displacement current, although there is no need to actually define any distinction between the two since all current is treated as simply generic motion of
All three effects in Figure 5.3 will be shown to have a major impact on the difference frequency susceptibility; self-detection provides a new and more viable route to low-frequency operation (<2 THz), while permanent dipoles also have a significant quantitative effect across the entire THz range. The atom-like susceptibilities are also still significant near their design resonances, at least for higher THz generation.

5.2.2 Steady-state equations

Given a Hamiltonian $H$ and density matrix $\rho$ of the periodic form expressed in Equations 2.40 2.41 2.42, the submatrices including the module energy difference are further decomposed into components at arbitrary frequencies label by $\alpha$:

$$H_p = \sum_\alpha H_p^{(\omega_\alpha)} e^{i\omega_\alpha t}$$

$$\rho_p = \sum_\alpha \rho_p^{(\omega_\alpha)} e^{i\omega_\alpha t}$$

$$\Delta_p = \sum_\alpha \Delta_p^{(\omega_\alpha)} e^{i\omega_\alpha t}.$$  \hspace{1cm} (5.3) (5.4) (5.5)

The optical field enters into the calculation in an electric dipole sense $[H^{(\omega_\alpha)} = qE(\omega_\alpha)z]$, where $E(\omega_\alpha)$ is the optical field and $z$ the position operator. Although this treatment destroys translation invariance in $H$, application of the Liouville-von Neumann equation will access only differences in the diagonals of $z$, ensuring translation invariance in the complete model. An important consequence of the electric dipole treatment, however, is that the module energy difference fluctuation in Eq. 5.5 is crucial in the treatment of optical nonlinearities.

The way that our model is able to calculate coherent and nonperturbative frequency mixing is seen by examination of the steady-state condition relating the different submatrices and frequency components (Eq. 2.49), restated here:
\[ i \omega_m \rho_p^{(\omega_m)} = \sum_{qn} \left( -\frac{i}{\hbar} \left[ H^{(\omega_m - \omega_n)}, \rho_q^{(\omega_n)} \right] - \delta_{pq} \Delta^{(\omega_m - \omega_n)} \rho_q^{(\omega_n)} \right) + \text{incoherent}. \] (5.6)

Solution of these equations is shown in Section A.2, including the inclusion of the incoherent scattering processes. The scattering processes are assumed to have only dc influence, so that they do not couple frequencies, which amounts to neglecting any ac changes in the scattering rate due to the bandstructure fluctuating in the presence of the optical fields. The coupling of different frequencies \( \omega_m \) and \( \omega_n \) is seen by 5.6 to occur through the Hamiltonian at their difference frequency \( \omega_m - \omega_n \).

### 5.2.3 Velocity extraction

Once the steady-state density matrix solution is found, it encodes all of the known information on the electronic system, and from it we can extract all the transport and optical properties including current, gain, and nonlinear susceptibility. Because we have an infinitely long chained system with periodic boundary conditions, the polarization is not a uniquely defined quantity, as it can in general depend on the boundary positions; therefore, all quantities must be derived from the velocity, which is uniquely defined. Supposing we have, in general, a time-evolution superoperator for the density matrix \( X \) \( (\cdot \rho = X \rho) \); the velocity expectation value is then

\[ \langle v \rangle = \text{Tr}(z \dot{\rho}) = \text{Tr}(z X \rho) = \sum_{ab,cd} X_{ab,cd} \rho_{cd}. \] (5.7)

In other words, we must evaluate the full sum with all possibilities of (density-matrix element at \( cd \)) \times (evolution from \( cd \) to \( ab \)) \times (position element at \( ab \)). We need a scheme adapted to our periodic system, so we invoke a requirement for convenience that the module be drawn in such a way that the intermodule dipole-matrix elements and transition rates are nonexistent, amounting to a mandate that the module boundary is drawn at the thick tunneling barrier. This location is not necessarily identifiable in all QCL systems, but it is
in our case and serves to simplify the mathematics. A visualization of one possible way to evaluate the sum under this assumption is in Figure 5.4. The starting and ending points of the arrows correspond to the $cd$ and $ab$ elements in Equation 5.7, respectively. Blue arrows depict terms which move entirely within the intramodule submatrix and red those which move into the intramodule submatrix from the outside. Under the assumption that there is no intermodule dipole operator or transitions, the illustrated combinations constitute the fully representative and nonredundant set.

Mathematical details on solving the steady-state equation are given in A.2, where it is shown how to retrieve the velocities at all frequencies included in the model. In simulating a system of two mid-IR pumps at frequencies $\omega_2$ and $\omega_3$ with terahertz difference frequency $\omega_1$, the parameters of interest (current density $J$, first-order susceptibility $\chi^{(1)}$, and second-order susceptibility $\chi^{(2)}$) can be inferred using Equations 2.50, 2.51, 2.52.

5.2.4 Example: Two-level system with permanent dipole

A mock system can be imagined which helps to isolate the effect of permanent dipoles on the difference frequency susceptibility. This is a two-level system, pictured in Figure 5.5, where the two states have dipole matrix element $z_{12}$ but also a difference in their centers $\Delta z$. Only a simple downward relaxation time of 1 ps will be included in the system dynamics, and the electron density is taken as $N_d = 10^{16} \text{ cm}^{-3}$.

When the diagonals of $z$ are not taken into account, the system has inversion symmetry, and so cannot possibly have a second-order (or any even-order) susceptibility. Therefore,
Figure 5.5: (Top) schematic of a two-level system with permanent dipole. (Bottom) Difference frequency susceptibility calculated at (left) a fixed signal of 10 meV, (right) a fixed pump of 100 meV.
conventional use of the SOS expression, which drops terms involving any diagonal of \( z \), gives \( \chi^{(2)} = 0 \). However, when we compute the full steady-state solution, some amount of \( \chi^{(2)} \) is found when the state centers are included. The left plot of Figure 5.5 shows generation of a fixed signal at \( \hbar \omega_1 = 10 \) meV, where the two pumps are swept keeping this constant difference. Peaks in the difference frequency susceptibility occur when one of the pumps is on resonance with the transition: this is due to the off-resonant pump modulating the susceptibility of the resonant pump by modulating the energy separation. SOS expressions can be calculated with the diagonals of \( z \) included, but this results in a translational variance. The condition \( z_{11} = 0 \) turns out to produce the correct answer for this simple two-level system, but in more complicated systems there can be no correct choice of origin.

The right side shows a scenario where one of the pumps \( \omega_2 \) is fixed at the transition resonance, while the other pump is swept, also sweeping the difference frequency. The peak in this scenario is at the optical rectification limit, when the pumps are degenerated to produce a difference frequency at dc. In this scenario the translational variance of the SOS expressions is especially pronounced. The optical rectification susceptibility of this system does in fact have a perturbative solution which was derived more carefully by Rosencher, et. al. [90]:

\[
\chi^{(2)}(0 = \omega - \omega) = \frac{16N_d q^3(\rho_{11} - \rho_{22})z_{12}^2 \Delta_z^2}{\epsilon_0 \hbar^2} \frac{\omega_{21}^2}{[\omega_{21} - \omega^2 + \Gamma^2][\omega_{12} + \omega^2 + \Gamma^2]},
\]

(5.8)

which matches exactly with the optical rectification susceptibility calculated using the steady-state solver.

### 5.2.5 Example: Three-level system

Another mock system comprising three levels can be devised to demonstrate the ability of our method to seamlessly capture both the permanent dipole and the resonant (atom-like) contributions to difference frequency susceptibility, as well as effects of electromagnetically
induced transparency. This system is based on an asymmetric well pair separated by a \( \delta \)-potential, so that there is a single ground state and two anticrossed upper states. The anticrossing resonance occurs when the width of the left well \( L_L \) is one half the width of the right well \( L_R \) so that the ground state of the left well mixes with the first-excited state of the right well. The strength of the \( \delta \)-potential directly tunes the anticrossing separation, which for the purposes here we will simply assume to be 10 meV. The system along with relevant calculated and assumed parameters are shown in Figure 5.6, displaying the frequencies at the resonance condition. Downward transition rates of 1 ps are included from each of the two anticrossed upper states 2 and 3 downwards to the lower state 1. The dipole matrix \( Z \) is calculated by diagonalizing within the basis of three relevant states originating from the individual wells (left well ground state, right well ground and first-excited states), assuming that coupling is negligible between the two ground states since they are far detuned in energy. Diagonals of \( Z \) are outlined in red for visibility.

As was done in the previous section, the difference frequency susceptibility is analyzed on the left for a fixed difference frequency of 10 meV (resonant with levels 2 and 3), and on the right for a fixed mid-IR pump at 90 meV (resonant with levels 1 and 2). Comparisons are made between the full solution and the SOS expression, both with and without the diagonals of \( Z \) included. In both the “fixed-signal” and “fixed-pump” scenarios, the SOS expression appears to yield the exact same result as the full solution, when diagonals of \( Z \) are not included; however, the diagonals of \( Z \) have a quite noticeable effect, especially closer to the optical rectification limit, and the SOS expression does not properly account for this.

### 5.2.6 Example: Electromagnetically-induced transparency

The three-level system in the previous section can be borrowed to demonstrate the model’s ability to capture electromagnetically-induced transparency (EIT). The level arrangement falls into the category of the \( \Lambda \)-configuration: in this scenario, EIT is possible if an ac pump field at resonance between states 2 and 3 is strong enough to produce distinct dressed states, which would in turn alter the absorption (linear susceptibility) of a weak probe ac field near
Figure 5.6: (Top) Schematic, assumed parameters, and calculated parameters of the three-level system created by coupling an antisymmetric well pair by a \( \delta \)-potential. (Bottom) Calculated difference frequency susceptibility with added pure dephasing \( T_2^* = 300 \) fs.
either the 1-2 or 1-3 resonance [175]. EIT is recognizeable, and named, specifically by a strong reduction in the absorption of the weak pump at line center. Ordinarily in EIT, the lifetimes of the upper states ($\tau_{21}$ and $\tau_{31}$) are far apart in magnitude, but for the purposes here we will still consider these times to be the same since they are both decay to the ground state; EIT effects are still visible.

The strong pump field is $\omega_1$, near the 2-3 resonance, and the probe field will be $\omega_3$, near the 1-3 resonance. An expression exists for first-order susceptibility seen by $\omega_3$ in the presence of the strong pump, which was derived using a perturbative approach to the probe field while treating the pump field nonperturbatively [172]:

$$\chi^{(1)}(\omega_3) = \frac{N z_{31}^2}{\epsilon_0 \hbar} \frac{\omega_1 - \omega_{21} + i/2\tau_{21}}{|\Omega|^2 - (\omega_3 - \omega_{31} + i/2\tau_{31})(\omega_1 - \omega_{21} + i/2\tau_{21})}.$$  \hspace{1cm} (5.9)

$\Omega = q|E(\omega_1)|z_{23}/\hbar$ is the pump field coupling strength to the 2-3 transition, where $|E(\omega_1)|$ is the pump field magnitude. An important quantity is $\Omega/\omega_{32}$: the definition of the ultra-strong coupling regime is when this becomes comparable to 1. This expression ignores pure dephasing, which is dropped here to make the EIT effect more recognizeable.

Figure 5.7 shows the results obtained for two different pump intensities $I_1 = 100, 1000$ W/mm$^2$, which result in $\Omega/\omega_{32} = 0.04, 0.11$. Diagonals of $Z$ are dropped here for the purposes of a comparison with the EIT expression (5.9). The emergence of the EIT effect in the steady-state solver is shown as more frequencies (labeled in the legend) are allowed in the movement of $\rho$ (see Eq. 5.4). When only the pump and probe frequencies are allowed, the EIT effect is not visible, but once the difference frequency between them ($\omega_2 = \omega_3 - \omega_1$) is included, the bulk of the EIT effect appears to be captured. A minor correction comes after the inclusion of the sum frequency $\omega_3 + \omega_1$, which is more important with stronger pump intensity; after this, additional corrections arising from further mixing with the pump are negligible. The full steady-state solution does not seem to converge exactly to the EIT expression, which could be because of the rotating-wave approximation made in the derivation of the EIT expression for both the pump and probe frequencies. The notable difference is an asymmetry between the two absorption peaks that occurs in the steady-state solution but not in the
Figure 5.7: Emergence of the EIT effect as more frequencies are included in the steady-state solution.

EIT expression. This example highlights an important point that while our method is non-perturbative within the included frequencies, care should be taken to include the relevant frequencies when analyzing nonlinear phenomena.

A much bigger correction to the absorption in this system under EIT is due to the permanent dipoles. Figure 5.8 shows the solutions calculated both with and without the diagonals of $Z$ included, for intensities varying on a logarithmic scale. As the pump intensity grows, the permanent dipoles begin to have a large effect on the probe absorption, which does not exist at all when the pump is turned off. This effect appears to be manifested in an asymmetry between the two split absorption peaks, and is significantly larger than the asymmetry introduced by the absence of a rotating-wave approximation shown in Figure 5.7. At the left side of the figure for the highest intensity, the side of the higher peak EIT arising from the split 1-2 transition is visible.
Figure 5.8: Effect of permanent dipoles on EIT, comparing the steady-state solution with and without diagonals of $Z$ included, for different intensity levels.

5.3 Scattering calculations

We now return to the concept of THz difference frequency generation in mid-IR QCLs and our model system. This section outlines the methods used to calculate transition and dephasing rates due to scattering, which are an input to the steady-state density matrix solver. The transition and dephasing rates used here are in contrast to the scattering treatment in the next chapter, which is more general; here, the single-module tight-binding basis explained in Sections 2.4.2 and 2.4.3 is used so that only transitions and dephasing are necessary. Calculations in this section follow the treatment used in [176], which are rooted in even more fundamental theory in [177] and [178]. Further reading on scattering calculations can be found in [102].

Some important variable definitions that will be used throughout this Section are:

\[ m, n = \text{subband indeces} \]
\[ E_{mn} = E_m - E_n \]
\[ \tau_{m\rightarrow n}(E) = \text{transition time } m \rightarrow n \]
\[ T^{*}_{2,mn}(E) = \text{pure dephasing time } m \leftrightarrow n \]
\[ \Theta(\tilde{E}) = \text{Step function of energy difference } \tilde{E} \]
We will be calculating the transition and dephasing times $\tau_{m \rightarrow n}(E)$ and $T^*_{2,mn}(E)$, which are functions of the in-plane energy $E$, for all of the considered scattering mechanisms: interface roughness, LO-phonons, alloy disorder, and ionized impurities. Specifically, $\tau_{m \rightarrow n}(E)$ is the transition rate out of in-plane energy $E$ in subband $m$ to all of subband $n$, and $T^*_{2,mn}(E)$ is the dephasing rate of the coherence between subbands $m$ and $n$ which is held at energy $E$. Once these rates are calculated resolved in $E$, they are averaged over the subband assuming a distribution. Details on this are in Section 5.3.5.

The pure dephasing actually arises from intrasubband transitions, although there is an interesting point to be made. Coming from Fermi’s Golden Rule and for elastic scatterers, the transition rate between a single state at in-plane energy $E(k)$ in subband $m$ and all states of in-plane energies $E(k')$ in subband $n$ is:

$$\frac{1}{\tau_{m \rightarrow n}(E)} = \frac{2\pi}{\hbar} \sum_{k'} |\langle nk'|H_{\text{scatt}}|mk\rangle|^2 \delta[E_{mn} + E(k) - E(k')],$$

(5.10)

where $H_{\text{scatt}}$ is the scattering Hamiltonian. This can be used directly for the intersubband transitions, where $n \neq m$, but for the dephasing (which comes from the intrasubband transitions), the expression to use is:

$$\frac{1}{T^*_{2,mn}(E)} = \frac{\pi}{\hbar} \sum_{k'} |\langle mk'|H_{\text{scatt}}|mk\rangle - \langle nk'|H_{\text{scatt}}|nk\rangle|^2 \delta[E(k) - E(k')].$$

(5.11)

The important point is that the two matrix elements of $H_{\text{scatt}}$ inside the $||$ are squared in a sense that the cross terms are kept. If the cross terms were dropped, this expression would be simply the lifetime dephasing coming from the intrasubband transitions in both subbands, but keeping the cross terms can cause a reduction in the dephasing. In fact, it can be seen later in the equations for each scattering mechanism that dephasing vanishes completely between states that have the same probability density, which happens for example in tunnel coupled states. Physically, this reduction in dephasing is a transfer of coherence: if an electron exists in a superposition of subbands at a definite wavevector, when it scatters...
to another wavevector the superposition can be preserved somewhat. When subbands have more similar probability density, their interaction with the scatterer is also more similar, making this effect important.

This Section will use two example transitions taken from Active Region A in Figure 5.1: a mid-IR transition between states 12-10, and a THz transition between states 11-10.

5.3.1 LO-phonons

LO-phonons are the only inelastic scattering mechanism considered in this modeling. In the equations, it is useful to use the following definitions:

\[
K = \frac{m^*e^2E_{LO}}{8\pi\epsilon_0\hbar^2}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_{dc}}\right) \tag{5.12}
\]

\[
F_{(ab)(cd)}(q) \equiv \int dz \int dz' \psi_a(z)\psi_b(z)\psi_c(z')\psi_d(z')e^{-\sqrt{q^2+q_s^2}|z-z'|} \tag{5.13}
\]

\[
q_s^2 = \frac{1}{L_d^2} = \frac{q^2N_d}{\epsilon_{dc}k_BT}. \tag{5.14}
\]

The wavevector \(q_s\) is the inverse of the Debye length \(L_d\), which introduces screening into the model. For an electron temperature of 500 K, \(L_d\) in our model system is approximately 26 nm.

The transition rates are given by:

\[
h\tau_{m\rightarrow n}^{-1}(E) = (n_{LO} + 1)C + n_{LO}D
\]

\[
C(E) = 2K\Theta(E + E_{mn} - E_{LO}) \int_0^\pi d\theta \frac{1}{q_e} F_{(mn)(mn)}(\tilde{q}_e)
\]

\[
D(E) = 2K \int_0^\pi \frac{1}{q_a} F_{(mn)(mn)}(\tilde{q}_a)
\] \tag{5.15}
\[
q_e^2 = 2k^2 + \frac{2m^*}{\hbar^2} (E_{mn} - E_{LO}) - 2k \sqrt{k^2 + \frac{2m^*}{\hbar^2} (E_{mn} - E_{LO}) \cos \theta}
\]

\[
\dot{q}_a^2 = 2k^2 + \frac{2m^*}{\hbar^2} (E_{mn} + E_{LO}) - 2k \sqrt{k^2 + \frac{2m^*}{\hbar^2} (E_{mn} + E_{LO}) \cos \theta},
\]

and the dephasing rates are:

\[
hT_{2, mn}^\ast (E)^{-1} = (n_{LO} + 1) A + n_{LO} B
\]

\[
A(E) = K\Theta(E - E_{LO}) \int_0^\pi d\theta \frac{q_e}{q_e} \left[ F_{(mn)(nn)}(q_e) - 2F_{(nn)(mn)}(q_e) + F_{(mm)(mm)}(q_e) \right]
\]

\[
B(E) = K \int_0^\pi d\theta \frac{q_a}{q_a} \left[ F_{(nn)(nn)}(q_a) - 2F_{(nn)(mn)}(q_a) + F_{(mm)(mm)}(q_a) \right]
\]

(5.16)

\[
q_e^2 = 2k^2 - \frac{2m^* E_{LO}}{\hbar^2} - 2k \sqrt{k^2 - \frac{2m^* E_{LO}}{\hbar^2} \cos \theta}
\]

\[
q_a^2 = 2k^2 + \frac{2m^* E_{LO}}{\hbar^2} - 2k \sqrt{k^2 + \frac{2m^* E_{LO}}{\hbar^2} \cos \theta}.
\]

Figure 5.9 shows the calculated scattering rates for our example transitions. The activation energy for LO-phonon emission is visible in the terahertz transition; below this energy only the slower LO-phonon absorption occurs. LO-phonon scattering is also fairly fast across the mid-IR transition, such that the upper state lifetime gets reduced to the sup-picosecond level once transitions to all subbands are accounted for. Dephasing due to LO-phonons also has a distinct shape in the in-plane energy, showing a visible activation of the emission process at \(E_{LO}\). The picosecond-level dephasing rates make LO-phonons an important dephasing, in addition to transition, mechanism.
5.3.2 Interface roughness

Modern growth technologies can produce quantum wells at near atomic-level precision, although there is always some amount of surface roughness. The roughness is random, and is usually described by parameters $\Delta$ and $\Lambda$, which are the average roughness height and in-plane correlation length, respectively. The scale of $\Lambda$ is important to how interface roughness should be viewed: if $\Lambda$ is much longer than the mean free path of electrons, it is best to think of separate electrons existing within different well widths, which is an inhomogeneous broadening, whereas if $\Lambda$ is shorter than the mean free path, the interface roughness is better thought of as a scattering mechanism causing homogeneous broadening [179]. Here, we will consider interface roughness as the latter.

The expression for transitions due to interface roughness scattering is:

$$
\hbar \tau_{m \rightarrow n}^{-1}(E) = \Theta(E + E_{mn}) \frac{m^*}{\hbar^2} \Delta^2 \Lambda^2 \delta U^2 \sum_i [\psi_m(z_i)\psi_n(z_i)]^2 \int_0^{\pi} d\theta e^{-\tilde{q}^2\Lambda^2/4}
$$

(5.17)

$$
\tilde{q}^2 = 2k^2 + \frac{2m^*E_{mn}}{\hbar^2} - 2k \sqrt{k^2 + \frac{2m^*E_{mn}}{\hbar^2} \cos \theta},
$$

and for pure dephasing we have:
Figure 5.10: Example interface roughness scattering times in Active Region A.

\[
\frac{\hbar T_{2,mn}^*}{2\hbar^2\Delta^2\Lambda^2\delta U^2} \sum_i \left[ |\psi_m(z_i)|^2 - |\psi_n(z_i)|^2 \right]^2 \int_0^\pi d\theta e^{-q^2\Lambda^2/4} \tag{5.18}
\]

\[q^2 = 2k^2(1 - \cos \theta).\]

\(\delta U\) is the step height of the roughness potential, or in other words the conduction band offset. The roughness parameters that gave reasonable agreement to experiment were \(\Delta = 0.8\) nm, \(\Lambda = 25\) nm. Results for the example transitions are in Fig. 5.10. Interface roughness is highly important as a dephasing mechanism, and contributes significantly to transport through the injector miniband, although we obtain negligible interface roughness transitions across the mid-IR transition for this long of a \(\Lambda\) which suppresses scattering of high momentum transfer. It should be understood that the interface roughness scattering parameters are the least precisely known inputs to our simulation, and in fact there is some debate as to whether an exponential roughness distribution might be more accurate than a Gaussian distribution [180]. Here we take the more traditional approach, which is to assume a Gaussian distribution and treat \(\Delta\) and \(\Lambda\) as fitting parameters [176, 102]. The equations can also be modified to include correlations between roughness in nearby interfaces [181], but this will be neglected here since the roughness parameters are used to fit anyways.
5.3.3 Alloy disorder

When materials are alloyed, the “alloy fraction” refers to the average concentration of a given species, which are distributed randomly. The usual assumption is that this distribution is completely random, with no correlation between the locations of any two atoms of the same species. Therefore, a random potential is introduced on the atomic scale which is treated as a scattering mechanism.

The equation for alloy disorder transitions in a ternary III-V semiconductor is:

\[
\hbar T_{m\rightarrow n}^{-1}(E) = \Theta(E + E_{mn}) \frac{1}{\hbar^2} \int dzm^*a^3(\delta E_c)^2x(1 - x) [\psi_m(z)\psi_n(z)]^2
\]

and for pure dephasing we have:

\[
\hbar T_{2,mm}^{-1} = \frac{1}{2\hbar^2} \int dzm^*a^3(\delta E_c)^2x(1 - x) [\psi_m(z)^2 - \psi_n(z)^2]^2
\]

\(a\) is the material lattice constant, \(x\) is the alloy fraction (of either alloyed atomic species), and \(\delta E_c\) is (in principle) the conduction band offset for the pure materials that are alloyed. However, it has been determined that often a different, smaller, \(\delta E_c\) should be used to better fit experimental data: we will follow [182] and use \(\delta E_c = 0.6\) eV for the In\(_{0.53}\)Ga\(_{0.47}\)As and 1.4 eV for the In\(_{0.52}\)Al\(_{0.48}\)As. The fractions \(x\) being close to 0.5 mean that alloy disorder scattering is close to maximized, as opposed to non-alloyed materials \((x = 0 \text{ or } 1)\) which of course exhibit no alloy disorder scattering.

Results for alloy disorder scattering in our example transitions are in Figure 5.11. Alloy disorder scattering exhibits no dependence on the scattering wavevector as seen in the equations. Both transitions and dephasing are relatively modest compared to the other mechanisms, even given the alloy fractions that are close to 0.5.
5.3.4 Ionized impurities

As QCLs are n-doped to produce conduction band electrons, the ionized dopants are left behind as scatterers. This is a non-local scattering mechanism governed by the Coloumb interaction, which is heavily reduced (but not eliminated) by screening. The same Debye screening length as was used for LO-phonons will be used here ($L_D = 26 \text{ nm}$). Further discussion on the validity of the Debye screening model can be found in [183]; a general rule is that the Debye model is valid when $L_D$ is on the order of or longer than the module length $L_{\text{mod}}$ (in our case $L_D = 26 \text{ nm}$ in comparison to $L_{\text{mod}} \approx 65 - 70 \text{ nm}$). The reason for this is that the Debye model assumes free electrons, and so for short screening lengths the corrections due to the subband energy structure become significant. A full treatment would require iterative calculation of the dopant potentials screened by the surrounding steady-state electron distribution.

The equation for the ionized impurity transitions is:

$$
\hbar \tau_{m \rightarrow n}^{-1} = \frac{m^* e^4}{4 \pi \epsilon_0^2 \hbar^2} \int_0^\pi \frac{d\theta}{\tilde{q}^2} \int dZ N(Z) \times \left\{ \int dz \psi_m(z) \psi_n(z) e^{-\tilde{q}|z-Z|} \right\}^2
$$

(5.19)

$$
\tilde{q}^2 = 2k^2 + \frac{2m^* E_{mn}}{\hbar^2} - 2k \sqrt{k^2 + \frac{2m^* E_{mn}}{\hbar^2} \cos \theta} + \frac{1}{L_d^2},
$$
and for the dephasing is:

\[
\hbar T_{2,mn}^*(E) = \frac{m^* e^4}{8\pi \epsilon_0^2 \hbar^2} \int_0^\pi \frac{d\theta}{q^2} \int dZ N(Z) \times \left\{ \int dz \left[ \psi_m(z)^2 - \psi_n(z)^2 \right] e^{-q|z-Z|} \right\}^2
\]

\[q^2 = 2k^2(1 - \cos \theta) + \frac{1}{L_d^2},\]

\(N(Z)\) is the 3D dopant density (shown as the shaded regions in Fig. 5.1). Calculated ionized impurity scattering times are in Fig. 5.12. Because of the steep drop-off of the ionized impurity scattering with wavevector, transitions are far faster across the THz transition than the mid-IR transition. However, dephasing due to ionized impurities is significant for both transitions, although more-so for the mid-IR one. The ionized impurity dephasing is closely centered around the subband minima, also owing to the steep drop-off with scattered wavevector.

### 5.3.5 Subband distributions

The subband filling statistics in QCLs under operating conditions have been an area of intense study, and are known to have a large impact on the overall transport characteristic. A laser is inherently a nonequilibrium device, and in a QCL equilibrium is broken in more
aspects than only the subband populations. Two additional effects are important: the subband electron temperatures $T_e$ tend to be significantly higher than the lattice (phonon) temperature $T_L$, and subband distributions can often be noticeably nonequilibrium, with hot electrons residing high in the subband, particularly in the lower lasing states [125, 184, 185]. In our model, we capture an approximation of both effects by assuming all subbands to be Boltzmann-distributed with $T_e = 500$ K, but with a certain amount of hot electrons superimposed as a Gaussian distribution at higher energy (centered at 140 meV above the subband minimum in light of elastic scattering across the radiative transition). The fraction of hot electrons is made highest in the lower lasing subbands (30%), decreases steadily to zero moving downstream through the injector, and remains at zero for the upper lasing states. This phenomenological scheme is designed to reflect the carrier distributions observed in detailed Monte Carlo simulations which resolve the in-plane k-states [125]. Including these nonequilibrium distributions are particularly important to obtain approximate quantitative agreement with experimentally observed current densities within the mid-IR lasers, which is central to the new mechanism of difference susceptibility that we identify.

The explicit equation for the subband distribution (fraction of population per unit energy) is then:

$$P_n(E_\parallel) = \frac{(1 - f_n)}{k_B T} e^{-E_\parallel/k_B T} + \frac{f_n}{\sqrt{2\pi}\sigma^2} e^{-\left(E_\parallel - \mu\right)^2/2\sigma^2},$$  \hspace{1cm} (5.20)

where $E_\parallel$ is the in-plane electron kinetic energy, $f_n$ is the hot electron fraction assigned to
the subband, $\mu = 140$ meV is the hot electron energy center, and $\sigma$ is the standard deviation (chosen as 25 meV). The hot electron fractions are assigned as the following:

$$f_n = \begin{cases} 
  f_{11} \frac{E_n - E_1}{E_{11} - E_1} & n \leq 11 \\
  0 & n \geq 12 
\end{cases}$$

(5.21)

where $f_{11} = 0.3$ is the hot electron fraction for the lower lasing state 11 and $E_n$ are the subband energies. A plot of selected subband distributions for Active Region B is shown in Fig. 5.13. Our choice of subband electron distribution is in light of other works, most notably the Monte Carlo simulation by Matyas et. al. (see [125] Fig. 3a), the NEGF results by Lindskog et. al.[184], and the experimental measurements by Spagnolo et. al. [185].

### 5.4 Results

We choose to model the active region from Vijayraghavan, et. al. of dual In$_{0.53}$Ga$_{0.47}$As / In$_{0.52}$Al$_{0.48}$As heterostructures [44] (bandstructures in Fig. 5.1). Although the two regions were designed for gain around 8.2 $\mu$m (37 THz) and 9.2 $\mu$m (33 THz), the transition linewidths were sufficiently broad that it was possible to achieve a large tuning range in the generated THz output from 1.7-5.25 THz by tuning the short wavelength pump in an external cavity setup. The device produced 120 $\mu$W of peak power at 4 THz using a dual-period DFG grating cavity, and approximately 15, 45, 15, and 5 $\mu$W for 5, 4, 3, and 2 THz, respectively, in the external cavity setup.

#### 5.4.1 Current and Gain

The two active regions are biased in series, and so must draw the same current, which in turn determines the possible bias combinations. Therefore, to choose a pair of biasing points, we must first simulate the transport characteristics. Bandstructure, tunnel couplings between all pairs of states, and scattering rates are computed at each bias to produce the characteristic shown in Fig. 5.14(a), where we can choose a pair of biases at current density 10 kA/cm$^2$. 
Figure 5.14: (a) Transport characteristics for Active Regions A and B. (b) Gain spectra and saturation for Active Regions A and B at bias points such that $J = 10 \text{kA/cm}^2$.

It is noted here that the transport characteristic levels off more than was experimentally observed; this is likely due to the fact that leakage to the continuum is not included in our model, which increases with the bias field. Therefore, although it appears in the model that it would be difficult to bias both active regions simultaneously at their highest gain point, in reality the leakage current helps to alleviate this constraint. Additionally, the optical field adds some extra amount of current which will be shown in the next Section.

Bandstructures calculated using a three-band k.p model at the chosen bias combination are shown in Fig. 5.1. Wavefunctions are calculated within a single module bounded by adjacent injection barriers, and tunnel couplings are calculated between all possible pairs of states in neighboring modules so as to include any possible injection channels (entering into $H_{1,1}$ of Eq. (2.40)). The tunnel couplings were calculated by direct evaluation of the k.p Hamiltonian matrix elements using all of the conduction, light-hole, and split-off wavefunction components which makes for a reliable scheme when nonparabolicity is significant. The exact method is outlined Appendix B.

Gain for the two active regions with increased intensity (each region treated independently) is shown in Fig. 5.14(b). These simulations include input at only a single frequency, and so neglect cross-saturation due to another pump. The longer wavelength active region exhibits less gain in the model than the shorter, because of the biasing condition explained above. The saturation intensity is realistic: 10 kW/mm$^2$ amounts to 2 W inside the waveguide with mode area 200 $\mu$m$^2$. 

112
5.4.2 Optical nonlinearity and photoconductivity

Rather than examine the difference frequency susceptibility \( \chi^{(2)} \) itself, we instead define a current susceptibility \( Y^{(2)} \), which is linked to the current, rather than polarization response:

\[
Y^{(2)}(\omega_1 = \omega_3 - \omega_2) \equiv \frac{J_{\omega_1}}{\varepsilon_0 E_{\omega_2} E_{\omega_3}} = i\omega_1 \chi^{(2)}(\omega_1 = \omega_3 - \omega_2),
\]

(5.22)

where \( J_{\omega_1} \) is the current density response at \( \omega_1 \) and \( E_{\omega_2,3} \) are the input electric fields at \( \omega_{2,3} \).

The actual THz power is then proportional to \( |Y^{(2)}|^2 \):

\[
P_1 = \frac{l_{coh}^2 |Y^{(2)}|^2 P_2 P_3}{8\varepsilon_0 c^3 n_1 n_2 n_3 S_{eff}},
\]

(5.23)

where \( l_{coh} \) is the coherence length, \( P_n \) are the powers inside the waveguide for each frequency \( \omega_n \), \( n_n \) are the refractive indices for the same, and \( S_{eff} \) is the effective area of interaction [92].

\( Y^{(2)} \) is a function of two independent variables, which we could choose to analyze over different lines, as long as the condition \( \omega_1 = \omega_3 - \omega_2 \) is retained. Figs. 5.15a,b show \( |Y^{(2)}| \) as a function of generated frequency \( \omega_1 \) with one pump \( \omega_2 \) fixed at energy 130 meV (9.5 \( \mu \)m). Pump frequency \( \omega_3 \) is thus swept in conjunction with \( \omega_1 \) for this scenario. Equal intensities are input in both pumps \( \omega_2, \omega_3 \), while the intensity in the generated frequency is assumed to be negligible. For both active regions, the resonant nonlinearities in the vicinity of 3-5 THz are visible, but they are added to a background which peaks at the optical rectification limit \( (\omega_1 = 0, \omega_3 = \omega_2) \). The nonzero value of \( Y^{(2)} \) at DC generation implies that a steady current is generated, rather than only a polarization; this is the root of the need to analyze \( Y^{(2)} \) since \( \chi^{(2)} \) exhibits a pole.

Insight into the mechanism behind this peak comes from a single-frequency simulation, shown in Figs. 5.15c,d. This simulation includes only one optical frequency, which is swept, tracking the increase in DC current \( \Delta J_{dc} \). In other words, the photoconductivity of the active region is captured. These functions are, not suprisingly, similar in shape to the gain profile, since the increase in current comes primarily from stimulated emission across the
Figure 5.15: (a,b) Simulated difference frequency current susceptibility $Y^{(2)}(\omega_1 = \omega_3 - \omega_2)$ with increasing pump intensity for (a) Active Region A (long wavelength) and (b) Active Region B (short wavelength). The difference frequency $\omega_1$ is swept with pump $\omega_2$ fixed at 130 meV and the other pump $\omega_3$ swept in conjunction. The black, blue, and green lines correspond to 0.1 kW/mm$^2$, 1 kW/mm$^2$, and 10 kW/mm$^2$, respectively, in each pump. Values at optical rectification are marked by the dots. (c,d) The increase in current as a function of a stimulating frequency as found from a single-frequency simulation for (c) the long wavelength and (d) the short wavelength active regions. (Input intensities are four times the intensity of the single fields in a,b.) Scaling by $2nc$ reproduces the value of $Y^{(2)}$ at optical rectification (denoted by dots), although only for small intensity. Equivalent points in the two simulations along the horizontal axis (single frequency input at 130 meV) are marked by the dashed lines, and resonant processes are denoted on each plot.
radiative transition. The value of $Y^{(2)}$ at optical rectification can be explained entirely by this mechanism, as shown by the dots connecting equivalent points in the two simulations. At least for vanishing intensity, we see that:

$$\lim_{\omega_3 \to \omega_2} Y^{(2)}(\omega_1 = \omega_3 - \omega_2) = 2nc \frac{\Delta J_{dc}(I_\omega)}{I_\omega},$$

(5.24)

with $n$ being the refractive index and $\Delta J_{dc}(I_\omega)$ the change in DC current due to intensity $I_\omega$ in single pump frequency $\omega$. For fair comparison, we choose $I_\omega = 4I_{\omega_2} = 4I_{\omega_3}$, which is the peak intensity when beating the two pumps. At higher intensity, the right side is found to underpredict the value of $Y^{(2)}$ at optical rectification, meaning that effects at fourth and even higher order begin to have importance. Specifically, this can be interpreted in terms of the harmonics of the beatnote itself; the left-hand side calculates only the first harmonic at $\omega_1$, while the right-hand side would give the complete peak-to-trough distance in the responding current. The fact that the latter underestimates the first means that the higher harmonics work to reduce this distance. Regardless, the quantity of interest is not the peak-to-trough distance but rather the first harmonic itself. Ability to account for this saturation effect highlights the advantage of the nonperturbative treatment used here.

The relationship between current and input intensity at pump energy 130 meV is shown in Figure 5.16, where the saturation effect, a nonlinearity in $Y^{(2)}$ itself, is clearly evident. The model predicts Active Region A to have less increase in current with intensity than Active Region B, which is an effect of the pump being further from the peak in current stimulation (approximately the same as peak in gain). This is seen also in Figure 5.15d as compared to Figure 5.15c, and is also reflected in the reduced height of the optical rectification peak in Figure 5.15b as compared to 5.15a.

5.4.3 Comparison to perturbative expressions

The nonlinear susceptibility in quantum well active regions has usually been estimated using a perturbative “sum-over-states” (SOS) expression, given as [172]:

115
Figure 5.16: Simulated increase in current density for both active regions in response to a pump at energy 130 meV. The intensity axis is extended to 40 kW/mm$^2$, the peak intensity when beating two pumps of 10 kW/mm$^2$ each. Linear expansion of each around zero intensity is given by the dashed lines.

\[ \chi^{(2)}(\omega_1 = \omega_3 - \omega_2) = \frac{N q^3}{\hbar^2 \epsilon_0} \sum_{lmn} z_{lm} z_{nm} z_{ml} (\rho_{kl}^{(0)} - \rho_{mm}^{(0)}) \]
\[ \times \left( \frac{1}{\omega_{nl} - \omega_1 - i\Gamma_{nl} + \omega_{nm} + \omega_1 + i\Gamma_{nm}} \right) \]
\[ \times \left( \frac{1}{\omega_{ml} + \omega_2 - i\Gamma_{ml} + \omega_{ml} - \omega_3 - i\Gamma_{ml}} \right). \]  

(5.25)

The triple sum over state indices $l, m, n$ is within a single module, with $\omega_{xy}$ being the resonant frequency between states $x$ and $y$, $z_{xy}$ the dipole matrix elements, $\rho_{xx}^{(0)}$ the populations at zeroth order (vanishing intensity) and $\Gamma_{xy}$ the decay rate of density matrix elements at $xy$. The SOS expression is not meant to handle permanent dipoles (diagonals of the $z$ operator), as it is intended for centrosymmetric atom-like systems, and it can be shown that introduction of these terms does not in general yield a translationally-invariant result. However, the permanent dipoles might in some cases provide a mechanism of intersubband second-order nonlinearity; one classic example is optical rectification in a two-state antisymmetric quantum well system where a case-specific expression had to be derived more carefully [90]. In systems with permanent dipole it is also conceivable that one pump could modulate...
Figure 5.17: Comparison of $Y^{(2)}(\omega_1 = \omega_3 - \omega_2)$ in Active Region B for the full calculation vs. a sum-over-states for three different scenarios: (a) swept-signal generation with one pump fixed at 130 meV, (b) fixed-signal generation of 4 THz, (c) fixed-signal generation of 1 THz. The magnitudes are shown in the main plots and phases are displayed in the insets. The calculations were done for vanishing intensity. Units for $|Y^{(2)}|$ are Mm/Vs.
the energy difference between spatially separated states (a first-order Stark effect). This modulates the first-order susceptibility seen by another pump close to resonance, so that \( \chi^{(1)}(\omega_2) \) could be modulated at \( \omega_3 \) or vice versa. The end result is a second-order nonlinearity only requiring resonance with respect to one of the pump frequencies. To test the hope that the translational variance is small, however, we will additionally consider the result of the SOS expression with permanent dipoles included to attempt to account for permanent dipole effects in the perturbative approach.

Figure 5.17 displays a comparison between the full calculation of \( Y^{(2)} \) and the SOS result both with and without permanent dipoles included in Active Region B this time analyzing: (a) as a function of generated frequency with one pump held at 130 meV, (b) as a function of pump frequency for generation of 4 THz, and (c) as a function of pump frequency for generation of 1 THz. Population inputs to the SOS expression are given from the steady-state solution itself for fair comparison. The phases of \( Y^{(2)} \) are given under the magnitudes, where for reference a phase of zero (positive real \( Y^{(2)} \)) implies velocity in phase with the beating of intensity. We find that while the SOS expression with permanent dipoles included provides a rough estimation of \( Y^{(2)} \) in the higher THz range (magnitude comparable and phase within \( \pi/8 \) at 4 THz), for frequencies in the lower THz the full calculation becomes absolutely necessary. Approaching optical rectification, both SOS expressions yield a vanishing \( Y^{(2)} \) since \( \chi^{(2)} \) is finite, meaning that any process describable using the sum-over-states has zero efficiency in that limit. It can also be seen from the stark difference in phases that even the SOS expression with permanent dipoles does not include all of the necessary processes; for generation of 1 THz the phase is off by approximately \( \pi/2 \), and at optical rectification the SOS expressions predict phase at opposite \( \pm \pi/2 \). This latter implies a DC polarization, when in fact \( Y^{(2)} \) has a phase of zero in that limit, corresponding to DC current.

5.5 Analysis

The effect that we have predicted can be described as the high-frequency tail of self-detection: the addition of two mid-IR waves amounts to a beating of intensity which stimulates current
response at the difference frequency. This is associated with the radiative transition, evident in the similar shape to the gain profile with respect to pump frequency (exhibited in all of Figs. 5.15c,d and 5.17b,c). Indeed, the increase of current with intensity is experimentally visible in QCLs, most notably as a discontinuity in the differential conductance at threshold as the onset of stimulated emission decreases the upper state lifetime. The response time associated with this mechanism is linked to the subpicosecond intersubband scattering times, which allows the bandwidth to reach into the THz range.

QCLs have an inherent design advantage in using this effect, since it is tied to the radiative transitions which the pumps are automatically close to resonance with. One might choose to approximate this detection effect by fitting to a simple response model:

\[ Y^{(2)}_{\text{detection}} \approx \frac{2nc}{1 + i\omega \tau} \beta, \]  

(5.26)

where \( \tau \) is a phenomenological response time and \( \beta \) is a coefficient for the current increase with intensity (\( \beta = \frac{\partial J}{\partial I} \) for vanishing \( I \)). At lasing intensity, however, the coefficient \( \beta \) is reduced because of the saturation of the detection effect, which is tied to gain saturation, exhibited in Figure 5.15. Nevertheless, we can fit \( \beta \) to the full model for different intensity levels. Figure 5.18 shows an approximate fit to this model, where the full calculation is compared to a simpler one where the detection \( Y^{(2)}_{\text{detection}} \) and SOS \( Y^{(2)}_{\text{SOS}} \) contributions are directly superimposed:

\[ Y^{(2)} \approx Y^{(2)}_{\text{detection}} + Y^{(2)}_{\text{SOS}}. \]  

(5.27)

Moderate quantitative agreement is found with the simple model, with \( Y^{(2)} \) being overpredicted by inclusion of permanent dipoles in the SOS expression but underpredicted by their exclusion. This suggests that permanent dipole effects even not linked to current beating play an important role reaching over the whole THz range, and that the translational variance of the SOS expression in accounting for them presents a significant error. Some error might also be introduced into the simpler model by the fact that \( \beta \) has frequency-
Figure 5.18: Fitting of the full calculation to a simpler model involving superposition of detection and SOS components. The black, red, and blue lines give results from the full calculation, SOS expression with permanent dipoles, and SOS expression without permanent dipoles. Calculations are performed on Active Region B, for the same fixed-pump scenario as in Figs. 5.15a,b and 5.17a. Real and imaginary parts are given for vanishing intensity in (a) and (b), and for 10 kW/mm$^2$ in each pump in (c) and (d). The fitted parameter $\beta = 0.110 ((kA/cm^2)/(kW/mm^2))$ for vanishing intensity, although at 10 kW/mm$^2$ in each pump this coefficient is reduced to 0.065 ((kA/cm$^2$)/(kW/mm$^2$)). The approximate response time $\tau$ is 240 fs.
dependence which would become important as the higher frequency pump moves further away, or by additional sources including the tunnel couplings which cannot be accounted for in the SOS expression.

Given that the current beating effect contributes significantly to difference frequency generation, it is useful to establish a way to estimate its strength in real devices using commonly measured experimental parameters. One such parameter is the differential conductance discontinuity at threshold $\Delta G$, and another is the “slope efficiency” of the output power $P_{out}$ vs. injection current $I_d$. Since the key parameter of interest is $\beta$, we solve for it at threshold:

$$\beta = \frac{\Delta G}{G_0 + \Delta G} \frac{1}{S} \frac{A_{\text{mode}}}{A_{\text{active}}} \frac{T}{T^2},$$

(5.28)

$G_0$ is the differential conductance just below threshold, $S$ is the slope efficiency defined as $dP_{out}/dI_d$, $A_{\text{mode}}$ and $A_{\text{active}}$ are the lasing mode and top-down active region areas, respectively, and $T$ is the output facet transmission (approximating that $T/2$ is the ratio of output power to total power inside the waveguide). Since $A_{\text{mode}}$, $A_{\text{active}}$, and $T$ are primarily cavity-related parameters, Eq. (5.28) provides a useful metric for comparison between different active regions by placement in the same cavity configuration. In a QCL, this expression would be approximately proportional to the population inversion [109], which is intuitively linked to the strength of current beating. It is important to note that the value of $\beta$ as found by (5.28) is for vanishing intensity and at threshold, but will likely still be indicative of the strength at more normal operating conditions.

The injection barrier thickness is well known to have large impact on the coherence of the injection process, and hence on $\Delta G$. Fig. 5.19 displays the effect of an altered barrier width on $Y^{(2)}$ for both active regions. The results suggest that there is some danger in suppression of the current beating by choosing too thick a barrier, and also that there may be some room for improvement by its reduction - at least in the case of Active Region A.

The collective results of this paper suggest a simple strategy for optimizing performance for low THz generation ($< 2$ THz). We suggest using only a single active region, which should have sufficient mid-IR gain bandwidth when the pumps are closer in frequency. Using
Figure 5.19: Effect of the injection barrier width (in legends) on $Y^{(2)}$ for vanishing intensity. Green lines correspond to values chosen in Ref. [44].

A single active region removes the current continuity constraint of using two active regions, and allows operation at the optimum point. Active region A appears to be preferred; that is to say, it is predicted to have a larger $Y^{(2)}$ value when biased at its optimum point, and has more room for improvement by thinning the injection barrier.

5.6 Conclusions

We have presented a density matrix transport model for QCLs which handles both gain and optical nonlinearity coherently and nonperturbatively. Scattering and dephasing processes were carefully accounted for by detailed calculation of the most relevant mechanisms, and care was taken to account for hot subband distributions and nonequilibrium electrons. The model predicts reasonable current levels as compared with experiment and predicts the expected amount of gain for lasing and a reasonable saturation intensity.

The computed nonlinearity exhibits a peak in efficiency, as opposed to merely susceptibility, at optical rectification, which is ascribed to the increase of DC current proportional to intensity. This increase is identified as a self-detection effect occurring through stimulated emission across the radiative transition, a familiar phenomenon in QCLs which has even seen recent use in THz phase-sensitive imaging systems [186, 187]. We have shown that the high-frequency tail of this effect extends into the THz range because of the subpicosec-
ond intersubband scattering times. The detection itself is highly nonlinear, meaning that fourth- and higher-orders are significant. A sum-over-states expression was found not to accurately reproduce the complete susceptibility, especially for generated frequencies in the lower THz range where the detection is strong. The complete susceptibility is reasonably well-matched by superposition of a fitted detection susceptibility and a sum-over-states expression, although even this is still not exact, since permanent dipoles and other mechanisms play a role in the nonlinearity which is not correctly accounted for by a sum-over-states. Finally, a metric was derived to assess the strength of self-detection for active regions through experimentally-accessible parameters, and suggestions were made for improvement of the performance in the lower THz range (< 2 THz).

The current beating, or self-detection, effect is large, and our results serve to explain the surprisingly low-frequency THz generation in a device demonstrated in the literature. The prediction of significant current susceptibility $Y^{(2)}$ extending to DC-generation suggests that the eventual low-frequency shutoff is owing more to other factors such as free-carrier absorption, phase matching, and output coupling, which are beyond the scope of this work. Regardless, a route forward to increased conversion efficiency in the lower THz range might aim to exploit this effect. There is no need to attempt to lower the frequency of a resonant nonlinearity whose conversion efficiency will scale downwards as the square of the generated frequency. Specifically, we have pointed out that a simple experimental metric useful for the assessment of such active regions is the relative differential conductance change at the onset of lasing.

Finally, it is possible that the formalism developed in this work is applicable to the study of the recently developed multi-mode QCLs and frequency combs [188, 189, 190]. Most directly, it is conceivable that the same detection process predicted here is responsible for the familiar radio-frequency beatnote generation, as a result of current beating from pairs of adjacent frequency lines. Further, a similar model could be extended to study the effect of radio-frequency modulation on such active regions [191, 192], particularly to assess the contribution coming from nonresonant transition energy modulation enabled by permanent dipoles. The formalism presented here could also be readily extended to encompass third
order nonlinearities, which would allow the study of comb generation itself.
CHAPTER 6

Density matrix transport with generalized scattering

To this point, the transport model put forward in different variations in this Thesis requires a phenomenological basis choice. This is necessary because the incoherent motion, or dissipation, of the density matrix is assumed to include only transitions and loss of coherence, or in other words coupling of populations to themselves and other populations, and coupling of coherences to themselves only. If the superoperator for incoherent evolution is $\mathcal{R} (\dot{\rho} = \mathcal{R}\rho)$, then this means that only the following categories of elements exist in $\mathcal{R}$:

$$
\begin{align*}
\mathcal{R}_{aa,aa} & \rightarrow \text{ lifetime decay of } a \\
\mathcal{R}_{aa,bb} & \rightarrow \text{ transition rate } b \text{ to } a \\
\mathcal{R}_{ab,ab} & \rightarrow \text{ dephasing rate of } ab,
\end{align*}
$$

(6.1)

for $a \neq b$. Under this assumption, it is necessary to decide a priori on a specific set of basis states within which transitions and dephasing occur. The intuitive choice is the energy eigenbasis of the heterostructure potential, but as explained in Section 2.4.2, this must usually be modified. The classic and simplest example is when states are evenly mixed across thick tunneling barriers, where a basis localized to either side of the barrier is phenomenologically chosen in attempt to capture the true effect of scattering, which is understood to occur between states more local than the heterostructure eigenstates. The choice of a localized basis is merely a device to cause the model to recover experimentally-observed effects of localization on the current density, by essentially forcing in localization based on a prior assumption of where exactly it occurs most strongly.
If a robust and reliable model applicable across all QCL designs is to be settled upon, it is necessary to eliminate the need for phenomenology and prior assumptions on part of the modeler. Also, in particular for many THz QCL designs which involve multiple tunnelings across a module, it is common for no obvious basis choice to even exist, and in fact it is not even clear whether any basis choice allows scattering to enter the model to its full character. For a concrete example of how no basis choice might suffice, we can consider again the case of states separated by a barrier, for the bases shown in Fig. 2.3: the delocalized basis would capture transitions between the states but not any dephasing, whereas the localized basis would capture dephasing but not any transitions. It is usually assumed that in this simple system the localized basis can be chosen with minimal adverse effect due to omission of the transition, but once more complicated systems are drawn, it becomes necessary to not have to attempt making these choices. In fact, it is for this reason that phenomenological density matrix models of THz QCLs have been able to treat only the simplest designs of 3-4 states, leaving many of the higher-performance designs accessible to only rate-equation or NEGF modeling.

This Chapter will demonstrate that it is possible to return to the unambiguous choice of the heterostructure eigenbasis, if scattering is calculated in a more general sense than only transitions and dephasing. From the viewpoint of this basis, the effect of localization by scattering (for example) implies that transitions involve not only transfer of population, but also influence on the coherences. This could be categorized as population-coherence coupling, and one can readily imagine that other new couplings are possible as well (coherence-population coupling, coherence-coherence coupling). In general, it appears that a full treatment of scattering should include not only transitions and dephasing, but rather more generally coupling between any density matrix element and any other, simply:

$$\mathcal{R}_{ab,cd} \rightarrow \text{generalized motion in } \rho,$$

for any $a,b,c,d$. This sort of generalized scattering superoperator will be derived, tailored for applicability to QCLs, shown to produce effects of localization, and applied to various THz
QCLs in this Chapter.

The modeling reported in this Chapter shares similarities with only a few previous studies in the literature on QCLs, although the fundamental theory, including the strategy for introducing irreversibility, is much like that of Kohn and Luttinger [193]. A generalized scattering superoperator, the main ingredient for this analysis, was presented in [194] and used to study transport in a superlattice, although only LO-phonon scattering was considered, and key ingredients including the velocity associated with incoherent motion were missed. In [195], the authors used a similar approach and equivalent superoperator to [194] to study dynamics in a THz QCL, leading to insights on interaction with an optical pulse. The intriguing conclusion was made that quantum coherences are necessary to recover effects in the regimes of both thin and thick barriers: for thick barriers, coherence is important to describe the holdup of charge behind barriers, whereas for thin barriers, coherence is important to describe the nonlinear optical dynamics. In [196], the same generalized scattering superoperator was used to study dynamics and in-plane electron distributions in a THz QCL in the absence of an optical field. The work in this Chapter will take a slightly different approach to the derivation of the superoperator, leading to a slightly different result. Then, direct steady-state solution will be used to study transport, gain, and electron distributions in- and out-of-plane, in the presence of a strong optical field.

The Chapter is organized as follows: Section 6.1 gives a derivation of the superoperator, for both elastic and inelastic scattering, culminating in a well-compacted form. Section 6.2 develops details of implementation necessary for use in a QCL: specifically, periodic boundary conditions, and accounting for the continuous density of electronic states. Section 6.3 outlines how a superoperator is constructed for each individual scattering mechanism. Section 6.4 applies the theory to a simple superlattice system to demonstrate specifically the capturing of localization in the model. Section 6.5 applies the theory to various THz QCL designs, demonstrating its versatility and providing insights into their operation. Concluding remarks are made in Section 6.6.
6.1 Generalized scattering superoperator

This section gives an informal derivation, from first principles, of the scattering superoperator for both elastic and inelastic scattering, leading to compact expressions of both. A more formal derivation will be given in [197].

6.1.1 Elastic scattering

We begin by stating that the density matrix evolves according to the Liouville-von Neumann equation, with the Hamiltonian separated into parts $H$ and $V$:

$$\dot{\rho} = \frac{1}{i\hbar}[H + V, \rho] = \mathcal{L}\rho + \mathcal{V}\rho. \quad (6.3)$$

$H$ is defined as driving the known evolution, which will be treated fully, while $V$ is the part of the Hamiltonian that is only known statistically and will be treated perturbatively (scattering). We are assuming here that there is only a single scattering “mechanism” fully described by $V$; generalization to multiple mechanisms and the appropriate definition of “mechanism” will be given in Section 6.1.3.

Liouville-type superoperators $\mathcal{L} \equiv [H, \ldots]/i\hbar$ and $\mathcal{V} \equiv [V, \ldots]/i\hbar$ are defined from $H$ and $V$. So far, Eq. 6.21 is in principle exact, albeit useless since the details of the Hamiltonian for the electron and all of its surrounding environment are neither known nor computationally tractable. Instead, we seek to introduce irreversibility into the evolution coming from $V$. To do so, an infinitesimal damping is added to the equation of motion for $\rho$, which will be useful later since it breaks time-reversal symmetry so that irreversibility can be introduced:

$$\dot{\rho} = \mathcal{L}\rho + \mathcal{V}\rho - \lim_{\gamma \to 0} \gamma \rho. \quad (6.4)$$

Next under a quasi-static approximation, setting $\dot{\rho} = 0$, we get the relations $\mathcal{V}\rho = -(\mathcal{L} - \gamma)\rho$, and also $\rho = -(\mathcal{L} - \gamma)^{-1}\mathcal{V}\rho$. Substituting these back into 6.4 leads to:
\[ \dot{\rho} = (\mathcal{L} - \gamma)\rho - \mathcal{V}(\mathcal{L} - \gamma)^{-1}\mathcal{V}\rho. \quad (6.5) \]

The first term is still identifiable as coherent evolution, and the second seems to be incoherent. The infinitesimal damping, or irreversibility, is now present in both terms, and we will choose to leave it out of the first but keep it in the second, leaving:

\[ \dot{\rho} = \mathcal{L}\rho - \mathcal{V}(\mathcal{L} - \gamma)^{-1}\mathcal{V}\rho. \quad (6.6) \]

Assuming that we are working in the energy eigenbasis, the \( \mathcal{L} \) superoperator is diagonal, specifically \( \mathcal{L}_{wx,yz} = \delta_{wy}\delta_{xz}(E_w - E_x)/i\hbar \). The elements of \( (\mathcal{L} - \gamma)^{-1} \) are then \( [\mathcal{L}^{-1}]_{wx,yz} = i\hbar\delta_{wy}\delta_{xz}(E_{wx} - i\epsilon)^{-1} \), where \( \epsilon = \hbar\gamma \). Following techniques in complex analysis, with the intention to integrate over energies later, the denominator can be substituted as [198]:

\[ \lim_{\epsilon \to 0} \frac{1}{E_{wx} + i\epsilon} \Rightarrow P \left( \frac{1}{E_{wx}} \right) - i\pi\delta(E_{wx}). \quad (6.7) \]

The first term denotes taking the principal value of the integral over energies, which is real, and the second term leads to the residue taken using the Dirac \( \delta \)-function, which is imaginary. Once incorporated, the second term will lead to a real superoperator, which is familiar as the incoherent evolution, while the first leads to an imaginary superoperator which is essentially a second-order correction to the coherent superoperator \( \mathcal{L} \). Since \( \mathcal{L} \) is constructed from the state energies, this principal value integral can therefore be viewed as a correction to the energy structure due to scattering. However, the typical assumption in a QCL is that such corrections are small in relation to the motion arising from \( H \) [137]. Neglecting the principal value, we arrive at elements \( [(\mathcal{L} - \gamma)^{-1}]_{wx,yz} = -\pi\hbar\delta_{wy}\delta_{xz}\delta(E_{wx}) \).

Given these, we can get the elements of the incoherent superoperator \( \mathcal{R} = -\mathcal{V}(\mathcal{L} - \gamma)^{-1}\mathcal{V} \).

Multiplying through, we get that:
\[ \rho_{ab} = \left[-\mathcal{V}(\mathcal{L} - \gamma)^{-1}\mathcal{V}\right]_{ab} = -\sum_{cdefgh} \mathcal{V}_{ab,ef}(\mathcal{L} - \gamma)^{-1}_{ef,gh}\mathcal{V}_{gh,cd}\rho_{cd} \]

\[ = \pi\hbar \sum_{cdef} \mathcal{V}_{ab,ef}\mathcal{V}_{ef,cd}\rho_{cd}\delta(E_{ef}). \quad (6.8) \]

Using the Liouville superoperator for \( \mathcal{V} \), we arrive at a specific equation for the elements of \( \mathcal{R} \) in terms of the scattering Hamiltonian:

\[ \mathcal{R}_{ab,cd} = -\pi\hbar \sum_{mn} \left( \delta_{bn}V_{am} - \delta_{am}V_{nb} \right) \left( \delta_{dn}V_{mc} - \delta_{cm}V_{dn} \right) \delta(E_{mn}). \quad (6.9) \]

Another form that the terms can be arranged in is:

\[ \frac{\hbar}{\pi} \mathcal{R}_{ab,cd} = V_{ac}V_{db} \left( \delta E_{ad} + \delta E_{bc} \right) - \sum_{m} \left( \delta_{ac}V_{dm}V_{mb}\delta E_{am} + \delta_{bd}V_{am}V_{mc}\delta E_{bm} \right). \quad (6.10) \]

In this form, the first two terms are identifiable as “in-scattering,” and the terms inside the summation as “out-scattering” [194]. Finally, we also note that the full \( \mathcal{R} \) itself can be written in the following form, which will be useful as a compact notation and in the derivation of inelastic scattering in the next Section:

\[ \mathcal{R}\rho = \frac{\pi}{\hbar}[\delta E \circ [V, \rho], V], \quad (6.11) \]

where \( \delta E \) holds a matrix of \( \delta \)-functions, defined such that \( [\delta E]_{mn} = \delta(E_{mn}). \) This form can actually be derived easily by using \( \mathcal{R} = -\mathcal{V}(\mathcal{L} - \gamma)^{-1}\mathcal{V} \) and \( \mathcal{V} = [V, \ldots]/i\hbar, \) \((\mathcal{L} - \gamma)^{-1} = -\pi\hbar\delta E \circ \ldots.\)
6.1.2 Inelastic scattering

The superoperator for inelastic scattering is derived by assuming a joint density matrix for the electron and its surrounding boson bath (for example, an LO-phonon mode), and taking a trace over the bath to arrive at an equation of motion for the reduced density matrix of only the electron. The bath is assumed to be at thermal equilibrium. To begin, we use the same $\mathcal{R} = -\mathcal{V}(\mathcal{L} - \gamma)^{-1}\mathcal{V}$ as from the previous section, only now in the context of a combined electron-bath system:

$$\mathcal{R}_{eq}\rho_{eq} = -\mathcal{V}(\mathcal{L} - \gamma)^{-1}\mathcal{V}\rho_{eq}$$

$$\mathcal{V} = [V_{eq}, \ldots] = [V^\dagger b^\dagger + V b, \ldots]$$

$$(\mathcal{L} - \gamma)^{-1} = \delta E \circ \cdots = \delta (E_e + E_q) \circ \cdots$$ (6.12)

The joint density matrix $\rho_{eq}$ has the reduced density matrices $\rho_e$ and $\rho_q$ for the electron and bath, respectively. $V = V^\dagger b^\dagger + V b$ is now the electron-bath interaction Hamiltonian, where $V^\dagger, V$ act on the electron and $b^\dagger, b$ are the bosonic creation and annihilation operators. The $(\mathcal{L} - \gamma)^{-1}$ now takes into account the total energy of both the electron ($E_e$) and the boson ($E_q$).

We can see immediately that the equation for $\mathcal{R}_{eq}$ will have terms involving $b^\dagger/b^\dagger$, $b^\dagger/b$, $b/b^\dagger$, and $b/b$. However, terms involving $b^\dagger/b^\dagger$ and $b/b$ will result in zero once a trace over the bath is taken, leaving only the two terms involving $b^\dagger/b$ and $b/b^\dagger$. We can then separate the evolution into these two parts, defining:

$$\mathcal{R}_{eq}^{(1)} \rho_{eq} \equiv -\frac{\pi}{\hbar} [V^\dagger b^\dagger, \delta (E_e + E_q) \circ [V b, \rho_{eq}]]$$ (6.13)

$$\mathcal{R}_{eq}^{(2)} \rho_{eq} \equiv -\frac{\pi}{\hbar} [V b, \delta (E_e + E_q) \circ [V^\dagger b^\dagger, \rho_{eq}]]$$ (6.14)

where $\mathcal{R}_{eq} = \mathcal{R}_{eq}^{(1)} + \mathcal{R}_{eq}^{(2)}$. The superoperators for evolution of the electron reduced density matrix are:

$$\mathcal{R}_{eq}^{(1)} \rho_{eq} \equiv -\frac{\pi}{\hbar} [V^\dagger b^\dagger, \delta (E_e + E_q) \circ [V b, \rho_{eq}]]$$ (6.13)

$$\mathcal{R}_{eq}^{(2)} \rho_{eq} \equiv -\frac{\pi}{\hbar} [V b, \delta (E_e + E_q) \circ [V^\dagger b^\dagger, \rho_{eq}]]$$ (6.14)
matrix is found by the procedure:

\[
\mathcal{R}^{(1)} \rho_e = \text{Tr} \left\{ \mathcal{R}^{(1)} \rho_{eq} \right\}_{\text{bath}} = -\frac{\pi}{\hbar} \text{Tr} \left\{ [V^\dagger b^\dagger, \delta(E_e + E_q) \circ [V b, \rho_{eq}]] \right\}_{\text{bath}} \tag{6.15}
\]

\[
\mathcal{R}^{(2)} \rho_e = \text{Tr} \left\{ \mathcal{R}^{(2)} \rho_{eq} \right\}_{\text{bath}} = -\frac{\pi}{\hbar} \text{Tr} \left\{ [V b, \delta(E_e + E_q) \circ [V^\dagger b^\dagger, \rho_{eq}]] \right\}_{\text{bath}}, \tag{6.16}
\]

where also \( \mathcal{R} = \mathcal{R}^{(1)} + \mathcal{R}^{(2)} \). Evaluating the commutators and \( \circ \) results in eight terms, for which the trace over the bath is taken for each using a similar procedure. As an example, the trace for one of the terms in \( \mathcal{R}^{(1)} \) is (ignoring the \(-\pi/\hbar\)):

\[
\text{Tr} \left\{ V^\dagger b^\dagger [\delta(E_e + E_q) \circ (V b \rho_{eq})] \right\}_{\text{bath}}
\]

\[
= \sum_a \left\{ V^\dagger b^\dagger [\delta(E_e + E_q) \circ (V b \rho_{eq})] \right\}_{aa}
\]

\[
= \sum_{ab} V^\dagger b_{ab} [\delta(E_e + E_q) \circ (V b_{bc} \rho_{eq,ca})]
\]

\[
= \sum_{abc} V^\dagger \delta_{b+1,c} \sqrt{a} [\delta(E_e + (b - a) \hbar \omega_q) \circ (V \delta_{b+1,c} \sqrt{c} \delta_{ac} n_a \rho_e)]
\]

\[
= \sum_a V^\dagger \sqrt{a} \delta(E_e - \hbar \omega_q) \circ (V \sqrt{a} n_a \rho_e)
\]

\[
= V^\dagger [\delta(E - \hbar \omega_q) \circ (V \rho_e)] N_q. \tag{6.17}
\]

The indices \( a, b, c \) are used to label the Fock states of the boson mode, with \( n_a \) being the occupation probability of \( a \) and \( N_q \) the Bose-Einstein factor. After the trace procedure is performed over all of the terms, the end superoperator is found to be:

\[
\mathcal{R} \rho_e = \frac{\pi}{\hbar} \sum_{\pm} \left( N_q + \frac{1}{2} \pm \frac{1}{2} \right) \times \left( [\delta(E \pm E_q) \circ (V^\dagger \rho_e), V^{-/}\rho_e] - [\delta(E \mp \hbar \omega_q) \circ (\rho_e V^{-/}), V^\dagger \rho_e] \right). \tag{6.18}
\]
The $V^{\dagger/-}$ represent $V^{\dagger}/V$. With the understanding that a summation of the superoperators over phonon modes will be made including both the positive and negative of any given mode $q$, and that $V_q = V_q^{\dagger}$, the $V^{\dagger}/V^-$ can be swapped in individual terms to compact the notation even further:

$$\mathcal{R} \rho_c = \frac{\pi}{\hbar} \sum_{\pm} \left( N_q + \frac{1}{2} \pm \frac{1}{2} \right) \times \left[ \delta(E \pm \hbar \omega_q) \circ (V^{\dagger}_c \rho_c) - \delta(E \mp \hbar \omega_q) \circ (\rho_c V^{\dagger}), V \right].$$  (6.19)

The + and − correspond to, respectively, emission and absorption processes. Importantly, both $\mathcal{R}^{(1)}$ and $\mathcal{R}^{(2)}$ contain both emission and absorption. $\mathcal{R}$ can also be expressed in an elementwise form:

$$\frac{\hbar}{\pi} \mathcal{R}_{ab,cd} = \sum_{\pm} \left( N_q + \frac{1}{2} \pm \frac{1}{2} \right) \times \left( V^{\dagger}_{ac} V_{db} \left[ \delta(E_{ad} \pm \hbar \omega_q) + \delta(E_{cb} \mp \hbar \omega_q) \right] 
- \sum_m \left[ \delta_{ac} V^{\dagger}_{dm} V_{mb} \delta(E_{am} \mp E_{LO}) + \delta_{bd} V^{\dagger}_{am} V_{mc} \delta(E_{mb} \pm \hbar \omega_q) \right] \right).$$  (6.20)

### 6.1.3 Combined

The equations derived so far are for a single scattering mechanism which has Hamiltonian $V$. Of course, an electron in a QCL is subjected to many scattering mechanisms at once, so the theory must be formulated to handle this. If we label these mechanisms with indices $m$, Eq. 6.21 becomes:

$$\dot{\rho} = \frac{1}{i\hbar} \left[ H + \sum_m V_{m, \rho} \right] = \mathcal{L} \rho + \sum_m \mathcal{V}_m \rho.$$  (6.21)

Now, the quasi-static condition is $\sum_n \mathcal{V}_n \rho = -(\mathcal{L} - \gamma) \rho$ and $\rho = -(\mathcal{L} - \gamma)^{-1} \sum_n \mathcal{V}_n \rho$, which are substituted to obtain:

133
\[ \mathcal{L}\rho - \sum_{mn} \mathcal{V}_n \mathcal{L}^{-1} \mathcal{V}_m \rho = 0. \quad (6.22) \]

The terms in the summation where \( m \neq n \) imply that two separate mechanisms act together to influence \( \rho \). However, such an effect would mean that a correlation exists between the two scattering mechanisms. We wish to be able to add in mechanisms which are uncorrelated, so these terms are dropped in a so-called “random phase approximation”, leading to:

\[ \mathcal{L}\rho - \sum_{m} \mathcal{V}_m \mathcal{L}^{-1} \mathcal{V}_m \rho = 0. \quad (6.23) \]

Now we can see that the superoperators for uncorrelated scattering mechanisms can be simply added. This includes not only different actual mechanisms such as LO-phonon scattering, impurity scattering, etc., but also uncorrelated scatterers within the same mechanism, for example impurities at different locations and different rough interfaces. Later, in Sec. 6.3, this justifies taking integrals over space where correlation in the scattering potential is assumed not to exist.

A neat form can now be given for the full equation of motion for \( \rho \) including sums over elastic mechanisms \( m \) and inelastic mechanisms (intended to be phonon modes) \( q \):

\[
\dot{\rho} = 0 = \frac{1}{i\hbar} [H, \rho] + \frac{\pi}{\hbar} \sum_{m} [\delta E \circ [V_m, \rho], V_m] \\
+ \frac{\pi}{\hbar} \sum_{q} \left( N_q + \frac{1}{2} \pm \frac{1}{2} \right) \left( [\delta (E \pm \hbar \omega_q) \circ (V_q^d \rho_e) - \delta (E \mp \hbar \omega_q) \circ (\rho_e V_q^d)], V_q \right). \quad (6.24)
\]
6.2 Implementation: periodicity and state continuua

Eq. 6.24 is in a generic form which is not yet directly useful for a QCL. Two further steps will be necessary: introduction of periodic boundary conditions and a means of extracting velocity under them, and accounting for the subband energy structure and continuous density of electronic states to rid the equations of the energy δ-functions. Until this latter point, the analysis will be kept general enough that subband indices and wavevectors need not be introduced into the state labeling. The last part of this section describes how the optical field enters into the calculation.

6.2.1 Periodicity

The sums in 6.24 are over scattering “mechanisms”, which are now understood to be uncorrelated scattering potentials. For now, the associated labels \((m \text{ and } q)\) will be dropped from \(V\), with the understanding that the end results are added. We assume a block-matrix form of \(V\) and \(\rho\):

\[
V = \begin{bmatrix}
\ddots & \vdots & \ddots \\
(V_{-1-1}) & (V_{-10}) & (V_{-11}) \\
\cdots & (V_{0-1}) & (V_{00}) & (V_{01}) & \cdots \\
(V_{-11}) & (V_{10}) & (V_{11}) \\
\ddots & \vdots & \ddots
\end{bmatrix}
\quad \rho = \begin{bmatrix}
\ddots & \vdots & \ddots \\
(\rho_{-1-1}) & (\rho_{-10}) & (\rho_{-11}) \\
\cdots & (\rho_{0-1}) & (\rho_{00}) & (\rho_{01}) & \cdots \\
(\rho_{-11}) & (\rho_{01}) & (\rho_{11}) \\
\ddots & \vdots & \ddots
\end{bmatrix},
\tag{6.25}
\]

and further the energy delta-functions \(\delta E\) and the position operator \(z\) will have a similar form. The formalism shown here can reach arbitrarily far from the diagonal, but will be truncated based on the energy selectivity and coherences spanning not more than one module (to be shown later). Periodic boundary conditions will be invoked to reduce the two-script notation to a single script \((\rho_{fg} = \rho_{g-f})\), but this can not be done for \(V\): since correlation is intended to be dropped between positions along \(z\) in the scattering potential,
the elements in $V$ are treated as functions of $z$ which are inner producted (not scalars), and so knowledge of the specific modules involved in $VV$ products must be retained. A concrete example is that $V$ might represent scattering from a single rough interface: $V_{01}$ and $V_{-10}$ will be different from each other because the states involved are in different locations relative to the particular interface.

First, a way to solve for the steady-state solution using the general evolution superoperator $\mathcal{R}$ will be shown. After this, the method to extract velocity, which has an important distinction, will be made clear.

Now in the periodic system, we define a notation incorporated into $\mathcal{R}$:

$$\mathcal{R}_{ab, cd}^{pq, rs} \rightarrow \text{influence from element } cd \text{ in submatrix } rs \text{ on element } ab \text{ in submatrix } pq.$$ 

The periodic boundary condition on $\rho$ is enforced by the single-script notation mentioned earlier that submatrix $\rho_{fg} = \rho_{g-f}$, so that now the unknowns are contained within the three submatrices $\rho_1, \rho_0,$ and $\rho_{-1}$, after truncating coherences spanning across more than one module. We seek a superoperator governing the motion of these matrices in specific, which takes the form:

$$\mathcal{R}_{ab, cd}^{p, rs} \rightarrow \text{influence from element } yz \text{ in all instances of submatrix } r \text{ on element } wx \text{ in any particular instance of submatrix } p.$$ 

Elements of this superoperator can be extracted as follows:
\[
\dot{\rho}_{ab}^p = \rho_{ab}^p = \sum_{rs} \sum_{cd} \mathcal{R}_{ab,cd}^{rs} \rho_{cd}^{rs} = \sum_{rs} \sum_{cd} \mathcal{R}_{ab,cd}^{rs} \rho_{cd}^{s-r} = \sum_{s} \sum_{cd} \left[ \sum_{r} \mathcal{R}_{ab,cd}^{rs} \right] \rho_{cd}^s
\]

\[
\therefore \mathcal{R}_{ab,cd}^{rs} \rho_{cd}^{s-r} = \sum_{r} \mathcal{R}_{ab,cd}^{rs} \rho_{cd}^{s-r+s}.
\]

(6.26)

The matrix of elements \( \mathcal{R}_{ab,cd}^{rs} \) will then lead to the steady-state solution in our system. To extract velocity, a distinctly different approach has to be taken, which is restated here from Eq. 2.56. In terms of simple state labels \( a, b, c, d \), we have that:

\[
\dot{\rho}_{ab} = \sum_{cd} \mathcal{R}_{ab,cd} \rho_{cd} = 0 \quad \rightarrow \quad \text{Steady-state equation}
\]

\[
v_{ab} = \sum_{cd} \mathcal{R}_{cd,ab} \dot{z}_{cd} \quad \rightarrow \quad \text{Velocity operator.} \quad (6.27)
\]

The important difference between the two is that the subscripts of \( \mathcal{R} \) are switched: for steady-state analysis we are concerned with evolution moving into the element \( ab \), whereas to extract velocity we are interested in evolution out of \( ab \). With this in mind, we can find the elements of the velocity operator, which also has a periodic submatrix form, as:

\[
v_{ab}^0 = \mathcal{R}_{ab,cd}^p \rho_{cd} = \sum_{rs} \sum_{cd} \mathcal{R}_{cd,ab}^{rs} \rho_{cd}^{rs} = \sum_{rs} \sum_{cd} \mathcal{R}_{cd,ab}^{rs} \left( \dot{z}_{cd} + \delta_{cd} \delta_{ab} L \right)
\]

\[
= \sum_{s} \sum_{cd} \left[ \sum_{r} \mathcal{R}_{cd,ab}^{rs} \right] \left( 1 + \frac{\delta_{cd} \delta_{ab} L}{z_{cc}} \right) \dot{z}_{cd}.
\]

(6.28)

where \( L \) is the spatial separation between modules. It is useful to define the inside of the brackets as another superoperator \( \mathcal{T} \), to find the velocity operator and has elements calculated in a way slightly modified from \( \mathcal{R}_{ab,cd}^{rs} \).
\[
v_{ab}^p = \sum_s \sum_{cd} T_{ab}^{p, s} T_{cd}^s = \sum_r R_{cd}^{p, 0} T_{ab}^{p, 0} \left( 1 + \frac{\delta_{ab}\delta_{cd}rL}{\Delta_{cc}} \right), \tag{6.29}
\]

where \( L \) is the spatial separation between modules.

So far, the superoperators \( R \) and \( T \) have been left general enough to hold in principle any type of evolution; now, we focus specifically on evolution from our generalized scattering superoperator. Using the elementwise equations (6.10 and 6.20), multiple algebra steps can be made including changes of variables to isolate in terms of submatrix difference \( r \) in the energy \( \delta \)-functions. This is aided by a trick for the \( VV \) products which is allowed by spatial periodicity:

\[
V_{ab}^{hj} V_{cd}^{kl} = V_{a}^{h+t} V_{b}^{j+t} V_{c}^{k+t} V_{d}^{l+t}.	ag{6.30}
\]

Below is for elastic scattering and still ignoring the \(-\pi/\hbar\) prefactor (inelastic scattering follows obviously):

\[
R_{ab}^{p, s} = \sum_r \left[ V_{ac}^{pp} V_{d}^{r+s} V_{cb}^{p+s} \delta E_{rd}^r + V_{ad}^{pp} V_{d}^{s+p} V_{cb}^{r+s} \delta E_{rc}^r \right. \\
\left. + V_{ac}^{pr} V_{d}^{r+s} V_{cb}^{p+s} \delta E_{rd}^r + V_{ac}^{pr} V_{d}^{s+p} V_{cb}^{r+s} \delta E_{rc}^r \right] \\
- \sum_m \left[ \delta_{ac} V_{dm}^{pp} V_{mb}^{r+s} \delta E_{am}^r + \delta_{ad} V_{dm}^{pp} V_{mb}^{r+s} \delta E_{am}^r \right. \\
\left. + \delta_{ad} V_{dm}^{pp} V_{mb}^{r+s} \delta E_{am}^r \right].
\]

The superscript \( r \) in the energy \( \delta \)-functions is a submatrix index, so specifically \( \delta E_{wx}^r = \delta(E_w - E_x - r\Delta) \), where \( \Delta \) is the successive module energy difference, calls for an energy difference between states that are \( r \) modules apart. This resonance can be satisfied for \( r \neq 0 \), for instance by resonance between the bottom of a subband and a high-\( k \) state in the next lower energy module, but truncation in \( r \) must be made at some point: here we restrict \(|r|\)
to at most 1.

The striking relation between the equations for $\mathcal{R}$ and $\mathcal{T}$ is that the $VV\delta E$ parts of $\mathcal{T}$ can be found from $\mathcal{R}$ simply by swapping $p_{ab} \leftrightarrow s_{cd}$. So, while $\mathcal{R}$ is being calculated, $\mathcal{T}$ can be calculated alongside it, only with the added corrections for module separation at each step.

### 6.2.2 State continuua

Until this point, the theory is formulated using only generic state labels $a, b, c, d$ within periodic submatrices. Now, we can focus specifically on sets of states within subbands by allowing these state labels to include wavevectors $k$ in addition to subband indices $a, b, c, d$. Also, since we have a continuum of states, the problem needs to be coarse-gridded.

A key assumption to make this analysis possible is that “cross-k” correlations, terms in the density matrix like $\rho_{ak,bk'}$ where $k \neq k'$, are negligible. An equivalent statement is that in-plane spatial correlations in the electron wavefunction vanish. Now, the elements of $\mathcal{R}$ that we need are:

$$
\mathcal{R}^{p}_{akbk} = \frac{\pi}{\hbar} \sum_{r} \left[ V_{s}^{r} r_{ak}^{s} V_{r}^{p+s} p_{bk}^{s} \delta(E - r\Delta)_{ak,dk'} \right.

+ V_{r}^{p} r_{ak}^{p} V_{r}^{s+p} s_{bk}^{p} \delta(E - r\Delta)_{ck',bk} \right.

- \sum_{mk''} \left( \delta_{ak,ck'} V_{s}^{r} r_{dk'}^{s} V_{r}^{p} p_{mk''}^{s} \delta(E - r\Delta)_{ak,mk''} \right)

+ \left. \delta_{mk',bk} V_{s}^{r+s} s_{mk''}^{r} V_{r}^{p+s} p_{ck}^{r} \delta(E - r\Delta)_{mk',bk} \right].
$$

(6.31)

Noting all of the $VV$ products above, we see that the associated momenta are always equal but opposite between the two produced $V$ elements. This is quite convenient to the analysis; it means that correlations between scatterers of different momenta are automatically dropped because the method never calls for them in the first place.
A wise way to perform the coarse-gridding is to take advantage of the cylindrical symmetry and the continuous density of states in parabolic subbands. With this in mind, and also the lack of cross-\( k \) correlations meaning all unknowns have single well-defined \( k \), grid points can be defined which hold equal numbers of density matrix elements by gridding \( k \) into equally-sized bins in wavevector direction \( \theta \) and in-plane energy \( E_k \). The width of these bins are, respectively, \( \Delta \theta \) and \( \Delta E \).

The unknowns we are looking for can then be explicitly defined as:

\[
\rho_{akbk} \equiv \sum_{\alpha} \rho_{ak_{\alpha}bk_{\alpha}} = S_k \rho_{ak_{\alpha}bk_{\alpha}} \tag{6.32}
\]

where \( \alpha \) are the labels for wavevectors inside the grid point \( k \) and \( S_k \) is the number of \( k \)-states in it. The assumption in the last part is that the density matrix elements are uniform within the group of \( k \)-states. The equation of motion for truly single states is:

\[
\dot{\rho}_{ak_{\alpha}bk_{\alpha}} = \sum_{s} \sum_{cd} \sum_{k'} \sum_{\beta} R_{ak_{\alpha}bk_{\alpha}}^{sck'_{\beta}dk'_{\beta}} \rho_{sck'_{\beta}dk'_{\beta}} \tag{6.33}
\]

Then we can sum this over the index \( \alpha \) and make the assumption of density matrix uniformity in each single term on the right side to relate the unknowns we are interested in. The superoperator we want is then:

\[
R_{akbk}^{sck'dk'} = \frac{1}{S_k'} \sum_{\alpha \beta} R_{ak_{\alpha}bk_{\alpha}}^{sck'_{\beta}dk'_{\beta}} \tag{6.34}
\]

The sum over the initial and final spaces will next be transformed to an integral. The procedure will be shown for the last term of elastic scattering (Eqn. 6.2.2), and the rest are simple to infer.

We are evaluating a sum over \( \alpha, \beta \):

\[
-\frac{1}{S_k'} \frac{\pi}{\hbar} \sum_{\alpha \beta} \delta_{bk_{\alpha}dk'_{\beta}} \sum_{mk''} \sum_{\gamma} V_{ak_{\alpha} mk''_{\gamma}}^{\gamma rsss} V_{mp''_{\gamma} mk''_{\gamma}}^{p+rss} \delta(E_{mk''_{\gamma}} - E_{bk_{\alpha}} - r\Delta) \tag{6.35}
\]
This term has a triple sum over actual states \((\alpha, \beta, \gamma)\), but one sum is eliminated by the kronecker \(\delta\). (Other terms that start as a double sum remain as a double sum and so are simpler.) Doing this and rearranging to see the sum over the two remaining spaces, we have:

\[
-\frac{1}{S_{k'}} \frac{\pi}{\hbar} \delta_{bd} \delta_{kk'} \sum_{r} \sum_{m_{k''}} \sum_{\alpha, \gamma} V_{r+s}^{+} V_{m_{k''}^+}^{+} V_{m_{k''}^-}^{+} \delta \left( E_{mk''}^+ - E_{bk}, -r\Delta \right)
\]  

(6.36)

Now we convert to an integral and make the assumption that the elements of \(V\) are constant within the two grid points:

\[
-\frac{1}{S_{k'}} \frac{\pi}{\hbar} \frac{1}{\Delta k^4} \delta_{bd} \delta_{kk'} \sum_{r} \sum_{m_{k''}} V_{r+s}^{+} V_{m_{k''}^+}^{+} \int d^2 \vec{k}_\alpha \int d^2 \vec{k}'' \delta \left( E_{mk''}^+ - r\Delta - E_{bk}, \right)
\]  

(6.37)

The integrals over \(\vec{k}, \vec{k}''\) can be converted to energy, giving us:

\[
-\frac{1}{S_{k'}} \frac{\pi}{\hbar} \frac{1}{\Delta k^4} \left( \frac{\Delta \theta m^*}{\hbar^2} \right)^2 \delta_{bd} \delta_{kk'} \sum_{r} \sum_{m_{k''}} V_{r+s}^{+} V_{m_{k''}^+}^{+} V_{m_{k''}^-}^{+} \delta \left( E_{mk''}^+ - r\Delta - E_{bk}, \right)
\]  

(6.38)

The operation \(O(\ldots, \ldots)\) is the overlap in energy for the two intervals. The constant out front can be more compactly expressed as (using the 2D spin-less density of states for \(S_{k'}\)):

\[
\frac{m^*}{4\pi \hbar^3} \frac{\Delta \theta}{\Delta E} A
\]

(6.39)

The factor \(A\) will be cancelled later by \(1/\sqrt{A}\) dependence in the \(V\) elements.

The rest of the terms follow in the same way (except simpler for the outscattering), and we get for elastic scattering:

141
\[ \mathcal{R}_{p_{ab}^k}^{s_{c'd'}k'}(s_{ck'}) = \frac{m^*}{4\pi\hbar^3} \frac{\Delta \theta}{\Delta E} A \sum_r \left\{ V_{s_{ak'}^r c'} V_{s_{dk'}^r}^r O(E_{ak'} - r\Delta, E_{dk'}) \right. \\
+ V_{s_{ak'}^r c'}^r V_{s_{dk'}^r}^r O(E_{ck'} - r\Delta, E_{bk'}) \right. \\
- \sum_{mk''} \left[ \delta_{bk',dk'} V_{s_{ak'}^r c'} V_{s_{mk''}^r}^r O(E_{mk''} - r\Delta, E_{bk'}) \right. \\
\quad \left. \delta_{ak',ck'} V_{s_{dk'}^r}^r V_{s_{mk''}^r}^r O(E_{ak'} - r\Delta, E_{mk''}) \right] \} \right. . \] (6.40)

And for the inelastic (with all \(VV\) products intended to be \(V^\dagger V\)):

\[ \mathcal{R}_{p_{ab}^k}^{s_{c'd'}k'}(s_{ck'}) = \frac{m^*}{4\pi\hbar^3} \frac{\Delta \theta}{\Delta E} A \sum_r \left\{ N_{\text{LO}} + \frac{1}{2} \pm \frac{1}{2} \right\} \times \\
\sum_{mk''} \left[ \delta_{bk',dk'} V_{s_{ak'}^r c'} V_{s_{mk''}^r}^r O(E_{mk''} - r\Delta \pm E_{\text{LO}}, E_{bk'}) \right. \\
\quad \left. \delta_{ak',ck'} V_{s_{dk'}^r}^r V_{s_{mk''}^r}^r O(E_{ak'} - r\Delta \mp E_{\text{LO}}, E_{mk''}) \right] \} \right. . \] (6.41)

The \(T\) superoperator used to extract velocity develops to account for the state continua by the same techniques as \(R\).

Finally, cylindrical symmetry can be used under this gridding scheme to greatly reduce the computation time. Since the density matrix is isotropic about the in-plane angles \(\theta\), the number of unknowns can be collapsed by summing over angle, explicitly now redefining the unknowns as:

\[ \rho_{p_{ab}^{k \theta}} = \sum_{\theta} \rho_{p_{ab}^{k \theta}}. \] (6.42)

The superoperator \(\mathcal{R}\) relating these is then:
\[ R^{p}_{abbk} \cdot c^{s}_{ck'dk'} = \sum_{\theta'} R^{p}_{abk0} \cdot c^{s}_{ck'\theta'dk'\theta'}. \quad (6.43) \]

and \( T \) is similarly:

\[ T^{p}_{abbk} \cdot c^{s}_{ck'dk'} = \sum_{\theta'} T^{p}_{abk0} \cdot c^{s}_{ck'\theta'dk'\theta'}. \quad (6.44) \]

### 6.2.3 Optical field

The inclusion of an optical field follows in a way similar to the other methods in this thesis. We generalize \( \rho \) and the entire time-evolution superoperator \( R \) (here intended to hold both coherent evolution and all scattering mechanisms) to now hold frequency dependence:

\[ \rho = \rho(0) + \rho(+) e^{i\omega t} + \rho(-) e^{-i\omega t} \]
\[ R = R(0) + R(+) e^{i\omega t} + R(-) e^{-i\omega t}. \quad (6.45) \]

The steady-state condition is no longer that \( \dot{\rho} = 0 \), but rather \( \dot{\rho} = \rho(0) + i\omega \rho(+) e^{i\omega t} - i\omega \rho(-) e^{-i\omega t} \). The system of steady-state equations then appears as:

\[
\begin{bmatrix}
R(0) & R(-) & 0 \\
R(+) & R(0) & R(-) \\
0 & R(+) & R(0)
\end{bmatrix}
\times
\begin{bmatrix}
\rho(-) \\
\rho(0) \\
\rho(+)
\end{bmatrix} =
\begin{bmatrix}
-i\omega \rho(-) \\
0 \\
+i\omega \rho(+)
\end{bmatrix},
\]

(6.46)

with the instances of \( \rho \) intended to appear as a vectorized list of the unknowns in the three representative submatrices and the superoperators \( R \) in corresponding matrix form. Our task is now to find the elements of \( R(\pm \omega) \) in the periodic system. Velocities at \( \pm \omega \) are straightforward: only the dc \( T \) is necessary, because the ac velocity operator \( v(\omega) = i[H(\omega), z]/\hbar \) vanishes for the optical Hamiltonian \( H(\omega) = qEz \).
6.3 Individual scattering mechanisms

For the elastic scattering mechanisms, a general procedure can be established to evaluate $VV$ products. Crucial to this is that we want to drop correlation in the scattering potential for both different points in $z$ and different points in-plane. Equivalent to the latter is that we need to drop correlation in different in-plane scattering wavevector, but this turns out to be automatic as it is evident in inspection of Eqs. 6.40 (and 6.41 for inelastic) that $V_{q_1}V_{q_2}$ is never called for $q_1 \neq -q_2$.

The general procedure to get the contribution of single terms coming from all $z$ is then by an integral (except for interface roughness, and LO-phonons which will be integrated over $q_z$):

$$V_{A k_1, B k_2}V_{C k_2, D k_1} \rightarrow \frac{1}{A} \int dz \tilde{V}_{A k_1, B k_2} \tilde{V}_{C k_2, D k_1} \tag{6.47}$$

$\tilde{V}$ is not totally meaningful by itself, but is rather just a tool to get to the right answer. The rest of this section will state the expressions to be used for the products $\tilde{V}_{A k_1, B k_2} \tilde{V}_{C k_2, D k_1}$, which are functions of $z$.

6.3.1 Alloy disorder

$$\tilde{V}_{X k_w, Y k_z} = \Delta \sqrt{a^3 x(1-x)} \psi_X(z) \psi_Y(z) \tag{6.48}$$

$\Delta = $ band offset parameter, $a =$ lattice constant, $x =$ alloy fraction.

6.3.2 Interface roughness

For the interface roughness scattering, a sum is performed over interfaces at locations $z_i$ rather than an integral over $z$:

$$V_{A k_1, B k_2}V_{C k_2, D k_1} \rightarrow \frac{1}{A} \sum_{z_i} \tilde{V}_{A k_1, B k_2} \tilde{V}_{C k_2, D k_1} \tag{6.49}$$
\[ \tilde{V}_{X_{k_w},Y_{k_z}} = \sqrt{f(|k_w - k_z|)}\delta E_c \psi_X(z_i)\psi_Y(z_i) \]  \hspace{1cm} (6.50)

\( \delta E_c \) is the band offset, and \( f(q) \) is the correlation function. A Gaussian correlation function is:

\[ f(q) = \pi h^2 l^2 \exp\left(\frac{-l^2 q^2}{4}\right), \]  \hspace{1cm} (6.51)

where \( h = \) average fluctuation height and \( l = \) correlation length.

### 6.3.3 Impurities

\[ \tilde{V}_{X_{k_w},Y_{k_z}} = N(z) \int dz' \tilde{V}_{|k_w - k_z|}(z - z')\psi_X(z')\psi_Y(z') \]  \hspace{1cm} (6.52)

\[ V_q(z'') = \frac{\epsilon^2 \exp(-|z''|\sqrt{\kappa^2 + q^2})}{2\epsilon \sqrt{\kappa^2 + q^2}} \]  \hspace{1cm} (6.53)

\( N(z) = \) impurity concentration, \( \kappa = \sqrt{ne^2/\epsilon dc k_B T} \) is the inverse Debye length.

### 6.3.4 LO-phonons

The LO-phonon contribution will be evaluated as an integral over \( q_z \) rather than \( z \). This is necessary because the interaction strength varies with \( \vec{q} \), so correlations in \( z \) should not be dropped. The electron part of the Fröhlich interaction Hamiltonian for a single mode \( \vec{q} \) is:

\[ V_{\vec{q}} = \frac{\alpha \epsilon^2 e^{i\vec{q}\cdot\vec{r}}}{\epsilon} \]  \hspace{1cm} (6.54)

We are looking for:

\[ \sum_{\vec{q}} \langle \Psi_{A_{\vec{k_1}}} | V_{\vec{q}}^\dagger | \Psi_{B_{\vec{k_2}}} \rangle \langle \Psi_{C_{\vec{k_3}}} | V_{\vec{q}} | \Psi_{D_{\vec{k_5}}} \rangle. \]  \hspace{1cm} (6.55)

Inserting the Frohlich interaction and performing the integral over the in-plane directions
gives:

\[
|\alpha|^2 \sum_{q_z} \frac{1}{q_z^2 + |\vec{k}_1 - \vec{k}_2|^2} \langle \psi_A | e^{-iq_z z} | \psi_B \rangle \langle \psi_C | e^{iq_z z} | \psi_D \rangle.
\]

Finally, we can convert to an integral and we have the contribution for LO-phonons:

\[
V_{Ak_1,Bk_2} V_{Ck_2,Dk_1} \to \frac{1}{A} \frac{E_{LO} e^{2(\epsilon^{-1} - \epsilon_{dc}^{-1})}}{4\pi} \int dq_z \frac{\langle \psi_A | e^{-iq_z z} | \psi_B \rangle \langle \psi_C | e^{iq_z z} | \psi_D \rangle}{q_z^2 + |\vec{k}_1 - \vec{k}_2|^2},
\]

which is used in place of the \(V_{Ak_1,Bk_2} V_{Ck_2,Dk_1}\) in Equation 6.41. It is noted that although the inside of the integral can be complex for a given \(q_z\), the value at \(-q_z\) is its complex conjugate so that the whole integral, and resulting superoperator, are guaranteed to be real. This allows the integral over \(q_z\) to be taken over only one direction and multiplied by two to reduce computation time. Also, for this reason, the \(e^{\pm iq_z z}\) can be safely swapped.

### 6.3.5 Omission of electron-electron scattering

The role of electron-electron (e-e) scattering in QCLs is largely unknown, and contrasting conclusions have been drawn in the literature as to its significance. However, it is expected that e-e scattering should be negligible in regards to intersubband transitions [122]. With regards to intrasubband transitions, e-e scattering can be expected to play a role in subband thermalization [102]; implementation of this is far simpler in Monte Carlo methods than using density matrices or NEGF. However, the role of e-e scattering as a dephasing mechanism is less clear: it appears that a view of e-e scattering times as a direct dephasing mechanism significantly overestimates the actual broadening, due to the preservation of subband coherence that can occur in scattering events [199]. This presents a difficulty that plagues all QCL transport models. Electron-electron scattering is neglected in the present treatment, perhaps at the cost of missing some thermalization or small dephasing effects. Recently, an improved model was introduced to approximate the effect of e-e scattering in NEGF calculations [204], which may be of some use in future density matrix calculations as
While the scattering superoperator holds a vast array of effects, the usual more simpler effects can still be identified within it. For example, starting with the inelastic scattering superoperator in Eq. 6.20, we find Fermi’s golden rule by setting $b = a$, and $d = c$, so that $\mathcal{R}_{ab,cd}$ refers to coupling of populations. Seeking emission from state $c$ to $a$, we obtain:

$$\mathcal{R}_{aa,cc} = \frac{2\pi}{\hbar} (N_q + 1) |V_{ac}|^2 \delta(E_{ca} - \hbar \omega_q),$$

(6.58)

which is precisely Fermi’s Golden Rule. Terms such as this can also be extracted from the final computed superoperator to reveal transition rates. Figure 6.1 shows this in comparison to a direct Fermi’s Golden Rule calculation, between example states 4 and 2 in Figure 6.4. Some complication is introduced by the periodic boundary conditions, but the intramodule phonon emission scattering can be isolated from the others based on the in-plane energy at

Figure 6.1: LO-phonon emission rates extracted from the coarse-gridded superoperator, compared with Fermi’s Golden Rule.
the destination.

The data points for the coarse-gridded result are placed at the center of the energy grid intervals. Similar validations were run for the other scattering mechanisms to assist in the development of the code.

6.4 Example: superlattice tunneling

To demonstrate that the model incorporates effects of localization by scattering, we consider a specific example: a superlattice biased such that the ground state of each well is at resonance with the first excited state of the next. If we take into account only these two levels, this is a system of two subbands per period. There are in fact other states in the superlattice, but we will neglect them here just for the sake of the demonstration.

The bandstructures we consider, in GaAs/Al$_{0.15}$Ga$_{0.85}$As, are shown in Fig. 6.2. The well length is held constant at 24.8 nm to give a separation of approximately 20 meV in the ground and first-excited states in each well, and barriers reaching from extremely thin (0.6 nm) to extremely thick (11.6 nm) are considered. Even though the anticrossing condition becomes highly sensitive to bias with thicker barrier, we can still always find the bias where the wavefunctions look approximately the same. The module energy drops are all close to 20 meV, but are adjusted slightly to account for the small Stark shifts that would otherwise move the states out of resonance.

Simulated current in this system, as compared to a semiclassical hopping picture, is shown in Figure 6.3. The lattice temperature is assumed to be 77 K, and the doping density was taken to be $N_d = 10^{16}$ cm$^{-3}$, distributed uniformly across the structure. For fair comparison, the semiclassical hopping picture is taken directly out of the full model by removing all coherences from the equations: this leaves a system of rate equations between populations which are solved in the same way.

The striking disparity between the full model and a semiclassical model is evident: whereas the full model predicts a strong decrease in the current with increasing barrier
Figure 6.2: Bandstructures in the superlattice system of two levels per period, with increasing barrier thickness but constant well width. Anticrossing biases are chosen. The highlighted states form one module.
Localization can also be quantified directly by taking the expectation value of the electron position. In the full superlattice, the electron density is equal in all wells by translational invariance, but the center of charge within a single module can be found using only $\text{Tr}(\rho_0 z_0)$,
where \( \rho_0 \) and \( z_0 \) are the intramodule density matrix and position operator (see 2.42 and 2.41). Using this, the electron is found to localize increasingly to the negative (upstream) side of the barrier as the width is increased, indicating a significant holdup of charge.

It is possible to point to specific terms in the generalized scattering superoperator that cause localization. For example, we can consider the effect that LO-phonon scattering has to induce coherence between anticrossed states coming from the populations of states above. Taking the superoperator in its more bare form from Eq. 6.20 (without periodic boundary conditions or continuous states), the element in \( \mathcal{R} \) that governs this is:

\[
\mathcal{R}_{ab,cc} = \frac{\pi}{\hbar} (N_{LO} + 1) V_{ac} V_{cb} \left[ \delta(E_{ca} - \hbar \omega) + \delta(E_{cb} - \hbar \omega) \right],
\]

where the coherence \( \rho_{ab} \) is intended to represent a coherence between the anticrossed subbands and \( \rho_{cc} \) is a population in the module above. The emission process is most relevant here. The key to localization lies in the \( V^\dagger V \) product, which we will consider for a single phonon mode \( q \):

\[
V_{ac} V_{cb} = \langle a | V_q^\dagger | c \rangle \langle c | V_q | b \rangle \rightarrow \propto \langle a | e^{-iq_z z} | c \rangle \langle c | e^{iq_z z} | b \rangle.
\]

We can see that for two states \( a \) and \( b \) of similar probability density, the sign of \( V_{ac} V_{cb} \) becomes determined by the signs of the basis wavefunctions \( a \) and \( b \) in the vicinity of the initial state \( c \): coherence is driven towards the sign of their product. This implies a constructive interference of \( a \) and \( b \) in the range of \( c \), which produces net localization behind the barrier because \( c \) does not extend beyond it.

### 6.5 Application to various QCLs

This Section demonstrates the versatility of the model by applying it to various THz QCL designs. For each, calculations are made of current, small-signal gain at different biases, large-signal gain at a single bias, spatial probability density at a single bias with increased
intensity, and subband distributions for both vanishing and nonvanishing intensity. Additionally, subband populations are given as well as some instances of a particularly telling parameter defined as $d_{ab}$:

$$d_{ab} \equiv \sum_k \rho_{akbk} z_{ba} + c.c.,$$  \hspace{1cm} (6.61)

which is a measure of the localization that is stored in the coherence between subbands $a$ and $b$. This is expected to be most significant between closely-spaced subbands, so the choice of which $d_{ab}$ elements to report is design-specific.

Some qualitative corrections may need to be made to the bandstructures due to the presence of space charge; in particular, space charge can be important to capture simultaneous alignment of both injection and extraction couplings. As it stands, the simulations presented in this Section are in the low-doping approximation where space charge is neglected.

### 6.5.1 Diagonal hybrid design at $\sim 3.4$ THz

Bandstructure and simulation results for the first structure, a diagonal hybrid design at $\sim 3.4$ THz, are shown in Fig. 6.4. The design uses a single injector well, followed by a diagonal radiative transition. Extraction occurs through the strongly coupled three-state miniband, after which LO-phonon emission occurs. The design is adapted from that of [34], which produced a record 1.01 W in pulsed mode at 10 K.

A small parasitic current is seen in the transport characteristic, after which the laser produces gain for a fairly large dynamic range. Negative differential resistance seems to begin around the bias of 50 mV/module, evident from both the transport curve and the sign of the DC gain — negative differential resistance is equivalent to gain at low frequency, and this is picked up because of the generality of the model. In fact, it appears that the sign of the DC gain gives a convenient method to infer the sign of the DC conductance, although any corrections due to modulations in the scattering superoperator are not accounted for. However, it appears that at least so far in this design, these corrections are minor enough
Figure 6.4: Bandstructure and simulated results for a diagonal hybrid design in GaAs/Al_{0.15}G_{0.85}As, aimed at $\sim$ 3.4 THz. The layer thicknesses in nm starting at the tunnel barrier are 5.1/10.3/1.7/10.7/3.7/8.8/3.7/17.2 and a sheet doping of $4.38 \times 10^{10}$ cm$^{-2}$ is assumed at the center of the wide well. The lattice temperature was set to 100 K.
that the sign of DC gain is a good indicator of electrical stability (negative/positive DC gain indicates stability/instability).

Effects of gain saturation are also calculated. The saturation intensity in this design is close to 1000 W/mm². The bottom left shows the spatial electron density as intensity at 3.4 THz is increased, aligned with the bandstructure above and color coded with the top right. Most interestingly, some holdup of charge behind the barrier is predicted at low intensity, but as intensity increases, the holdup becomes stronger. This means that the coherence between states 4 and 5 increases with intensity; this is due to destruction of the population inversion followed by fast extraction through the miniband and subsequent LO-phonon emission to settle in the next module but upstream of the barrier (the injector state in a localized basis). This may suggest that even thinner barriers could help to provide stronger injection to push the intensity even higher.

The bottom middle and right plots show the subband distributions without (middle) and with (right) strong 1000 W/mm² excitation at 3.4 THz. Total subband populations $\rho_{aa}$ and relevant $d_{ab}$ parameters are also given. At zero optical intensity, the subband distributions can be significantly nonthermal, although the optical field seems to be a powerful thermalizing force. A similar behavior has been predicted by Monte Carlo simulations [125]. We also notice that while the lower states are operating quite cold with low intensity, they heat significantly when the optical field is turned on. It should be emphasized here, however, that electron-electron scattering, which should contribute to subband thermalization, is not accounted for.

The $d_{ab}$ parameters quantify the offset in charge that is stored in a particular coherence; the quantity $d_{45}$ is found to be the strongest, which belongs to the two tunnel-coupled states forming the injector and upper state. Some upstream offset of -1.5 nm is seen at low intensity, and at high intensity this increases to -3.3 nm. This is the same effect that was visible in the shifting probability density with intensity (bottom left). Other $d_{ab}$ parameters within the extraction miniband are much smaller, although a small increase of $d_{23} = -0.1$ nm to -0.4 nm occurs with the addition of the optical field. The small values of these $d_{ab}$ elements suggest that extraction is highly efficient in this design. Importantly, all observed $d_{ab}$ values
are negative, implying upstream localization as expected.

6.5.2 Vertical design at ∼ 3.0 THz

The next design is from Williams, et. al. (2005), which notably operated up to maximum operating temperatures of 164 K in pulsed mode and 117 K in CW [200]. A variation of this design at 4.4 THz also held the record for highest CW power for over a decade [35]. Bandstructure and simulation results are in Fig. 6.5. The design employs a two-well injector to a vertical transition aimed at ∼ 3.0 THz, followed by resonant extraction and then resonant LO-phonon emission.
This design is predicted to also have a relatively large dynamic range, and this is corroborated by the experiments [200]. This is likely due to the large injector miniband width and distance from the lower states, allowing efficient injection over a large bias range.

At the selected bias of 52 mV, the bandstructure appears such that the main upper states are only states 4 and 5, and 3 is only beginning to contribute. State 3 holds a large share of the population at this bias, but would not likely add as much to the gain because of its weak overlap with the lower states 1 and 2. Nevertheless, the simulations predict that transport is still efficient enough via states 4 and 5 to produce lasing. The experiments in [200] seem to show an effect of two different slope efficiencies, where an increase in slope efficiency with bias was observed well past threshold for many of the reported temperatures; this may be due to the use of a two-well injector where two separate resonance conditions produce gain above threshold. The existence of gain over such a large bias range seems to suggest that injection is reasonably efficient in this design. Additionally, the high saturation intensity is predicted.

The spatial electron density seems to hardly change with optical intensity. This may suggest that there is room for improvement in the extraction, since an efficient extraction scheme would be expected to move charge out of the active wells after stimulated emission. The smallness of the parameter $d_{12}$ does not suggest efficient extraction coupling at this bias, because the basis states 1 and 2 are already delocalized; rather, the higher population of state 1 over state 2 indicates the holdup of charge before being extracted. Truly, the inclusion of space charge will be necessary to examine the strength of extraction in this design, because the dopants in the wide well will tend to bring states 1 and 2 closer to anticrossing. The large population of state 3 is also clear in the electron density.

The subband distributions with zero optical intensity appear to be even more nonthermal than in the case of the previous design. However, the optical field is once again predicted to have a powerful thermalizing effect. The parameters $d_{34}$ and $d_{45}$ within the injector miniband are small but not insignificant, indicating some amount of charge holdup that again increases with the optical field. Importantly, both transitions and localization appear to play a role in transport through the injector miniband, which would make this design unamenable to a
density matrix model without generalized scattering.

6.5.3 Vertical design at $\sim 2.1 \text{ THz}$

The third simulated design is adapted from the second to lower the operating frequency from $\sim 3.0 \text{ THz}$ to $\sim 2.1 \text{ THz}$. This device was demonstrated in [201], where lasing occurred up to 72 K in pulsed mode and 40 K in CW, with 1.2 mW of power in CW at 17 K.

In agreement with the experiments, this laser is predicted to have a smaller dynamic range compared with the previous two designs. This is primarily because the lower operating
frequency places the parasitic resonance at a bias closer to the design bias; it is seen in both
the transport characteristic and gain that at 40 mV the parasitic channel is still strong and
reduces the gain to near zero, and at 48 mV the device is into NDR. Even within the small
dynamic range, the gain is significantly lower than the 3 THz design (gain in the 20 cm$^{-1}$s
at $T_L = 77$ K as opposed to 40 cm$^{-1}$s at $T_L = 100$ K). Additionally, the values and shape of
the gain curve that we predict are close to the experimental estimations taken on a similar
design in [202].

The spatial probability density is predicted to vary little with optical intensity, as was
the case with its 3.0 THz counterpart. The bandstructure at the chosen bias of 45 mV is also
qualitatively similar to 52 mV in the 3.0 THz design; subband 3 is holding a large amount
of population but not contributing to gain. The lower state populations and the difference
between states 1 and 2 actually seem to be somewhat smaller, possibly indicating a more
efficient extraction. Space charge will likely have a similar effect on this design, however.
The coherences in the injector miniband, particularly $d_{45}$, appear significant, and are again
increasing with optical intensity as was observed in the diagonal design. The suggestion of
stronger injection coupling in that design may therefore apply here, although space charge
should be taken into account to make a real conclusion.

Finally, highly nonthermal subbands are again predicted without the presence of the
optical field, and the optical field is predicted to have strong thermalizing effect.

6.5.4 Scattering-assisted design at $\sim$ 1.8 THz

The final simulation results are from a design which was originally intended to act as a higher-
frequency 4 THz device similar to the diagonal design of Section 6.5.1, but it was found that
the device exhibited a second positive differential resistance beyond the intended injection
anticrossing, on which lasing occurred at the much lower 1.8 THz resonance between states
inside the intended extractor [203]. It is notable that in this bias range the design operated
at an anomalously high temperature of 163 K for such a low frequency. The complexity of
this device provides a unique testing ground to analyze what the model can actually capture.
Figure 6.7: Bandstructure and simulated results for a diagonal $\sim$ 4.0 THz/scattering-assisted $\sim$ 1.8 THz design in GaAs/Al$_{0.15}$Ga$_{0.85}$As. The layer thicknesses in nm starting at the tunnel barrier are 4.2/8.5/2.3/9.6/3.4/7.3/4.0/15.3 and a sheet doping of $3 \times 10^{10}$ cm$^{-2}$ is assumed at the center of the wide well. The lattice temperature was set to 100 K.
Some of the post-injection PDR is visible in the transport characteristic, although it is perhaps not quite as strong as was seen in the data. This PDR is due to the activation of LO-phonon scattering between states 4 and 3/2/1 with increasing bias, occurring because of the proximity of this transition to $E_{LO}$, a result of the original intent at a high-frequency design. This effect is opposed by the holdup of transport occurring as the tunneling injection is misaligned, but under the right conditions, can overcome it. As bias increases beyond the injection point, the emergence of gain at lower frequencies is evident, although another gain peak at 1 THz appears at possibly a higher level than the 1.8 THz peak. However, it should be kept in mind that a prediction of the actual lasing frequency should take into account the cavity losses, which could be stronger at 1 THz and the higher 4.5 THz, allowing lasing at 1.8 THz to occur.

Importantly, this example aids to point out a subtle effect that is left out of the present treatment. While the second PDR is predicted, no gain curves were found on or around it which exhibit negative gain at DC bias. We expect that this is due to the neglect of effects arising from the AC modulation of the scattering superoperator itself; whereas the model gives a neat way of assessing the response in current to small-signal AC fluctuation in the coherent part of $\mathcal{R}$, any accompanying fluctuations in the incoherent part are not present. This discrepancy is brought out in this case where the fluctuation of the incoherent part is important due to the bias position amidst activating LO-phonon scattering.

Finally, although the subbands in this design seem to be somewhat thermalized by the optical field, it appears less so than in other designs. This could be linked to the irregular operation that was observed in the experiments.

### 6.6 Conclusions

This Chapter has developed, from first principles, a density matrix model of transport and optical properties in QCLs that does not require any phenomenological basis choice. The key ingredient is a generalized scattering superoperator which couples any density matrix element to any other, so that the energy eigenbasis of the heterostructure potential can
safely be used. It is these more general terms that specifically cause effects of localization by scattering that would otherwise not emerge from a semiclassical treatment using the energy eigenbasis.

The key tools for use in QCLs were also developed, including periodic boundary conditions, electron state continua, and inclusion of an optical field. The model was applied to a representative superlattice system to demonstrate the capturing of localization effects through a decrease in current with thickening barriers. Finally, the model was applied to three separate QCL designs to demonstrate its versatility. Key features of the model’s predictions were corroborated with published results, and insights were found into experimentally-inaccessible aspects of operation including spatial electron densities, subband distributions, and scattering-induced coherences.

Overall, this work provides a density matrix description of transport in QCLs that automatically includes a phenomenon that is well known but difficult to quantify: electron localization by scattering. Bandstructures in QCLs are found to vary with bias in a complex fashion, but because of localization, the transport properties turn out to be smoother than expected from a semiclassical viewpoint of only hopping transitions. Much of this smoothness is recovered by our model, although in fact the experimental results tend to appear even smoother. This seems to suggest that further localization or thermalization effects are occurring in reality, quite possibly due to electron-electron scattering, which is neglected in the present treatment. However, it should be noted that approximations to electron-electron scattering in NEGF calculations can also yield false parasitic currents in certain designs [204]. Even further work could include the effects of modulation of the scattering superoperator by the optical field, although as it stands, the model is ready to serve a design tool across the THz QCL repertoire. The treatment of the optical field could also be generalized in the spirit of Chapter 5 to allow simulation of nonlinear optics, for example frequency combs and RF modulation.
CHAPTER 7

Conclusions

This Thesis has presented extensive theoretical testing of novel concepts for high temperature operation of THz QCLs. The core of the theory is a transport model based upon the density matrix, placing the complexity at a level between that of simpler rate equation models and more thorough Nonequilibrium Green’s Function (NEGF) models. The density matrix approach has key advantages over rate equations in that effects of coherence are properly accounted for, and advantages over NEGF because of its relative numerical lightness. This enables, for example, simple inclusion of an optical field, so that nonperturbative optical response and nonlinearities can be calculated easily, in contrast to NEGF where the transport problem is alone so intensive that treatment of the optical field is limited [140]. A related advantage exploited in Chapters 3 and 4 is the ability to accommodate an electron-phonon product Hilbert space in the case that the two interact strongly, which is not a practically-accessible concept in NEGF. Furthermore, the reduced computational burden should allow for density matrix models to be used more conveniently as a design tool when many repeated simulations are desired.

Density matrix descriptions are not in any way new to the QCL community; rather, this work has expanded the concepts to enable broader application, and in doing so uncovered important points about operation of QCLs that are useful to the experimental efforts but not accessible without theory. The introductory density matrix models for QCLs were based upon algebraic solutions in small systems, which serve mainly to produce useful analytic equations that give insight into the dependence of various performance metrics on design parameters. For example, the famous Kazarinov-Suris tunneling formula was derived from density matrix analysis in the first proposal for a QCL [128], and became a fundamental
tenet of the field to understand tunneling injection. More complicated analytic equations have since been derived to describe transport in specific systems of not more than 3-4 states [134, 129, 130], but the complexity of the algebra quickly leads to equations which are intractable, do not have such obvious utility, and are only narrowly applicable. The departure made in this work is to abandon the analytical approach in favor of full numerical simulation; instead of relying on equations as a guide, the designer can use physical intuition and employ the model as a tool to test ideas. Similar tools have been described few times in the literature [131, 195, 196]. This work serves to provide firm formulations that expand the bounds of density matrix modeling in QCLs, to the benefit of QCL theorists, and in doing so also identify key aspects of operation in novel QCL systems, to the benefit of the experimentalists.

New developments in theory include:

- Provided various ways to solve for the steady-state of the density matrix equation of motion, given partially coherent and partially incoherent evolution. This included a way for the optical field to enter the calculation coherently and nonperturbatively, enabling description of QCLs under saturated operating conditions (throughout).

- Presentation of the physics of electron-LO-phonon interaction in quantum dots, including how to treat the case of multiple phonon-coupled electronic transitions. This enables the use of an electron-phonon product basis to account for polaron formation in terahertz quantum dot cascade lasers (Chapters 3, 4).

- A framework to introduce generalized frequency content in the optical field so as to recover nonlinear optical effects. This led to the identification of various effects which are captured by this modeling but are absent from less general treatments (Chapter 5).

- Derivation of a generalized scattering superoperator from first principles and discussion of its inner workings. This was also developed explicitly to the point of applicability in QCLs, which require periodic boundary conditions. The generalized scattering treatment eliminates the need for awkward phenomenological basis choices used in other
density matrix models (Chapter 6).

New insights that aid the experimental efforts include:

- Identification of complications inherent in the scaling of QCLs to the quantum dot limit. Specifically, highly-split gain was predicted, which may not necessarily preclude lasing, but extreme subthreshold parasitic currents were predicted that would prevent operation (Chapter 3).

- Outline of a counterintuitive design strategy to avoid the expected problem of parasitic current in quantum dot cascade lasers, based on the insights gained from Chapter 3. The strategy withstood theoretical testing, including with a moderate level of quantum dot size inhomogeneity (Chapter 4).

- Identification of previously underappreciated mechanisms contributing to the THz difference frequency susceptibility in mid-IR QCLs. The most notable is an effect named “self-detection,” which is an increase in the conduction current through the QCL that follows the mid-IR intensity beatnote. This would be the most useful mechanism of difference frequency susceptibility for the generation of frequencies below \( \sim 2 \text{ THz} \).

- Development of a simulation tool that is sufficiently general for design across the THz QCL repertoire, and discussion of operational aspects of QCLs that are not accessible to experiment, such as subband distributions with and without an optical field.

The results of this Thesis also suggest a number of future research directions. With regards to a THz quantum dot cascade laser (QDCL), a limitation so far to the modeling has been that only the regime of smallest pillar diameter could be considered. While this is conceivably possible to reach in certain materials (for example \( \sim 30 \text{ nm diameter in InAs} \)), a model for use as the diameter is scaled downwards would ideally be able to handle larger diameters, requiring the inclusion of more lateral states. The approach described in Chapters 3 and 4 would become intractable for such a problem, making a more approximated approach necessary. One method might be to fully diagonalize the electron-phonon Hamiltonian, after
which the phonon decay and generation could be included as an intricate set of Fermi’s Golden Rule transitions between the different polaron states. However, care should be taken to determine whether such an approximation is valid in the face of dephasing and localization. Additionally, a means of accounting for electron-electron interaction (charging) should also be developed if the goal is to reach accurate prediction of experiment.

Still, perhaps the more pressing need at the current time in the development of a QDCL is experimental data on nanopillar heterostructures as an optoelectronic material. Evidence of bound states in nanopillars has been found via electrical measurements [81], but a first proof-of-concept necessary for a laser will be to witness an optical signature of an intraband transition. A means of accomplishing this through a photoconductivity spectral measurement on large-area nanopillar arrays is proposed in Appendix C.

The modeling results on THz difference frequency generation suggest routes forward on the experimental front. In particular, it can be expected that the effect of self-detection has the potential to produce significant power in the lower THz range. It seems that, currently, even though some generation of these lower frequencies is evident in the data, the power is highly inhibited by the larger cavity losses as compared to higher frequencies. Given that existing active regions already hold significant power in this range, it is worth investigating new concepts for waveguide design that might be better optimized for extraction in the lower THz. It is also worthwhile to employ the active region modeling technique for design optimization of the nonlinear susceptibility, although care should be taken not to compromise the mid-IR pump performance. Furthermore, the concepts used to study frequency-mixing in the context of difference-frequency generation can be further generalized to include other nonlinear optical phenomena in QCLs, including four-wave mixing and RF modulation in QCL frequency combs.

The generalized density matrix theory presented in Chapter 6 is near applicability for real design optimization of THz QCL active regions. Furthermore, the model could be extended in the spirit of Chapter 5 to study nonlinear optical phenomena, of which four-wave mixing and RF modulation are again an interesting direction of study. Still, further improvements to the theory might include an incorporation of space charge, electron-electron scattering,
and even possibly effects of the scattering superoperator modulating via the optical field. The latter point is likely to be important if effects of RF modulation are of interest, although no precursor yet exists in the QCL literature.
APPENDIX A

Steady-state density matrix solution and velocity extraction

This Appendix gives methods to include incoherent velocity, solve the steady-state density matrix, and extract velocity: an elementwise method in A.1 which was used in Chapter 2 and 3, and a method based on vectorization of the density matrix in A.2 which was used in Chapter 5. The method of Chapter 6 was contained there.

Both will begin from the steady-state equation of Eq. 2.49, restated here:

\[ i\omega_m \rho_{p}^{(\omega_m)} = \sum_{q\ell} \left( -\frac{i}{\hbar} \left[ H_{p-q}^{(\omega_m-\omega_n)}, \rho_{q}^{(\omega_n)} \right] - \delta_{pq} q \Delta^{(\omega_m-\omega_n)} \rho_{q}^{(\omega_n)} \right) + \text{incoherent}. \]  

(A.1)

A.1 Elementwise

In this case, we will consider only one optical frequency, and label the frequencies \( \omega, 0, \) and \( -\omega \) as 1, 0, and \(-1\). First for the incoherent evolution, an equation can be obtained at each element in A.1 by invoking the Liouville superoperator \( \mathcal{L} \) and a change of variables:

\[ i\omega_m \rho_{p,ab}^{(m)} = -\frac{i}{\hbar} \left( \sum_{q\ellcd} \left( \mathcal{L}_{p,q}^{(m-n)} \rho_{q,cd}^{(n)} \right) - pE_{mod} \rho_{p,ab}^{(m)} \right). \]  

(A.2)

Terms outside the quadruple sum can then be brought inside using Kronecker \( \delta \)-functions, yielding a system of equations providing the relation \( \sum_{q\ellcd} M_{ab}^{p,(cd)} \rho_{q,cd}^{(n)} = 0 \), where:
\[ M_{(ab)p',(cd)q}^{m,n} = -\frac{i}{\hbar} L_{abcd}^{(m-n)} + i \delta_{pq} \delta_{mn} \delta_{ad} \delta_{bd} \left( p \frac{E_{\text{mod}}}{\hbar} - m \omega \right) + S_{(ab)p',(cd)q}^{m,n} \]  

(A.3)

\( S_{(ab)p',(cd)q}^{m,n} \) is the incoherent contribution. The incoherent evolution is separated into that due to each transition and pure dephasing:

\[ \frac{d}{dt} \rho^{\text{inc}} = \sum_X L^X \rho + D \rho. \]  

(A.4)

\( L^X \) is the Lindblad superoperator for transition \( X \), which is constructed in the form: [144, 205]

\[ L^X \rho = C^X \rho C^{X\dagger} - \frac{1}{2} \left( C^{X\dagger} C^X \rho + \rho C^{X\dagger} C^X \right) , \]  

(A.5)

where \( C^X \) is the jump operator which induces the transition. For a simple transition \( \psi_i \rightarrow \psi_f \) having rate \( \Gamma_{i \rightarrow f} \), the associated jump operator is \( C = \sqrt{\Gamma_{i \rightarrow f}} |\psi_f\rangle \langle \psi_i| \). In this case, \( C \) will have only one nonzero element, but in a combined Hilbert space this may not be true; \( C \) itself must be expanded as a tensor product and thus can acquire more than one nonzero element, in which case transfers of coherence can occur.

For example, we can examine the collapse operator which is due to the transition of phonon mode \( N \) from \( n_N = 1 \) to \( n_N = 0 \). Given the allowed mode occupations, the collapse operator in the space of \( \{|n_N,n_{\alpha}\}\} \) is then \( C^{1N \rightarrow 0N} = \sqrt{\Gamma_{1 \rightarrow 0}} (|00\rangle \langle 10| + |01\rangle \langle 11|) \). If \( N_{el} \) electron degrees of freedom are included, \( C^{1N \rightarrow 0N} \) is further expanded to \( \sqrt{\Gamma_{1 \rightarrow 0}} \mathbf{1}_{N_{el}} \otimes (|00\rangle \langle 10| + |01\rangle \langle 11|) \). We will neglect correlations in the different transition processes between number states of a given phonon mode by including separate jump operators for each.

Once the jump operators are obtained, we need to use them to fill out elements of \( S_{(ab)p',(cd)q}^{m,n} \) in the chain-coupled system. These are defined as:
\[ [L^X \rho]_{p,ab}^{(m)} \equiv \sum_{qcd} S_{(ab)_{p,cd}^{(m)}}^{X\rho_{q,cd}} \]  \tag{A.6}

or in other words the coefficients relating variable \( \rho_{q,cd}^{(m)} \) to the evolution of \( \rho_{p,ab}^{(m)} \) due to transition \( X \). Importantly, we first distinguish between transitions which are correlated between modules and those that are not. In the former, the jump operator itself assumes a chain-coupled form which forms a single Lindblad superoperator (shown in A.7), whereas in the latter there exists a series of jump operators which form separate Lindblad superoperators which are then superimposed (shown in A.8).

\[
L \begin{pmatrix}
\ddots \\
(C) \\
(C) \\
\ddots 
\end{pmatrix}
\]  \tag{A.7}

\[
\cdots + L \begin{pmatrix}
\ddots \\
(C) \\
(0) \\
\ddots 
\end{pmatrix} + L \begin{pmatrix}
\ddots \\
(0) \\
(C) \\
\ddots 
\end{pmatrix} + \cdots \]  \tag{A.8}

In the first case, we find that the elements in \( S^X \) are:

\[
S_{(ab)_{p,cd}^{(m)}}^{X\rho_{q,cd}} = \delta_{pq} \delta_{mn} \left[ C_{ac} \bar{C}^\dagger_{db} - \frac{1}{2} \left( \delta_{bd} [\bar{C}^\dagger \bar{C}]_{ac} + \delta_{ac} [\bar{C}^\dagger \bar{C}]_{db} \right) \right], \]  \tag{A.9}

assuming that \( \bar{C} \) resides in the diagonal submatrices, and for the second we find that:

\[
S_{(ab)_{p,cd}^{(m)}}^{X\rho_{q,cd}} = \delta_{pq} \delta_{mn} \left[ \delta_{pd} C_{ac} \bar{C}^\dagger_{db} - \frac{1}{2} \left( \delta_{bd} [\bar{C}^\dagger \bar{C}]_{ac} + \delta_{ac} [\bar{C}^\dagger \bar{C}]_{db} \right) \right], \]  \tag{A.10}

independent of any displacement of \( \bar{C} \) from the diagonal. Since the collapse operator \( \bar{C} \) is always positive, we notice in the solution for \( S^X \) that correlated transitions can transfer
intermodule coherence while uncorrelated transitions do not.

The pure dephasing contribution $D$ is trivial. For a pure dephasing time $T_2^*$ applied to all coherences it is:

$$D_{(ab)p^q,(cd)r^s} = -\frac{1}{T_2^*}\delta_{pq}\delta_{mn}\delta_{ac}\delta_{bd}(1 - \delta_{p0}\delta_{ab}), \quad (A.11)$$

but can also easily be generalized to incorporate different dephasing times.

Finally, we must derive the expression for the incoherent contribution to velocity, and thus $J^{inc}$. We are interested in the expectation value of velocity due to incoherent processes, and so we equate

$$\langle v^{inc} \rangle \equiv \frac{d}{d} Tr (\rho z) |^{inc} = Tr \left[ \sum X L^X \rho z + D \rho z \right]. \quad (A.12)$$

By the assumed form of $z$ in (2.41), we have for both types of transitions that

$$Tr \left( L^X \rho z \right) = Tr \left[ L \left( \bar{C}^X \right) \rho_0 z_0 \right], \quad (A.13)$$

although this will only truly hold for transitions which do not cross the module boundary. It is possible, however, to extend so as to include those that do. The pure dephasing part is:

$$Tr \left( D \rho z \right) = -\frac{1}{T_2^*} Tr \left[ \left( \rho_0^{(0)} - \text{diag}\rho_0^{(0)} \right) z_0 \right], \quad (A.14)$$

leading to the incoherent contribution to the current: $J^{inc} = N_d q \langle v^{inc} \rangle$.

A.2 Density matrix vectorization

Another method to solve for the steady-state density matrix makes use of the vectorization operator, which is columnwise conversion of a matrix to a vector. In this method we will
consider the density matrix evolution as split into three different groups due to coherent evolution, transitions, and dephasing:

\[
\dot{\rho} = \dot{\rho}_{\text{coh}} + \dot{\rho}_{\text{trans}} + \dot{\rho}_{\text{deph}} = \sum_n i \omega_n \rho^{(\omega_n)} e^{i \omega_n t}.
\] (A.15)

Vectorization has the useful property that \( \text{vec}\{AB\} = (1_N \otimes A)\text{vec}\{B\} = (B^T \otimes 1_N)\text{vec}\{A\} \), for multiplication of two square matrices \( A \) and \( B \) each having dimension \( N \). Vectorization of Eq. (2.49) gives:

\[
\text{vec}\left\{ \dot{\rho}_{p\text{coh}}^{(\omega_m)} \right\} = e^{\omega_m t} \sum_{qn} \left[ -\frac{i}{\hbar} \left( 1_N \otimes H_p^{(\omega_m - \omega_n)} - H_p^{(\omega_m - \omega_n),T} \otimes 1_N \right) \right] \ldots
\] (A.16)

\[
-\delta_{pq} q E_{\text{mod}}^{(\omega_m - \omega_n)} \text{vec}\left\{ \rho_{p\text{coh}}^{(\omega_n)} \right\},
\] (A.17)

with \( E_{\text{mod}}^{(\omega_m - \omega_n)} \) as the scalar module energy at the different frequencies, including DC.

Next we move to the transition contribution, where we will work under the assumption that transitions only occur within the module. Each transition process has separate instances inside each module, each of which has its own Lindblad superoperator (as in Eq. (A5) in Ref. [132]) formed with a lone instance of the intramodule submatrix \( \bar{C}_X \) at the module position. The result of each can be found using a similar block matrix multiplication tactic as was used for the coherent contribution, and added to yield:

\[
\dot{\rho}_{p\text{trans}} = \sum_X \delta_{\rho0} \bar{C}_X^\dagger \rho_p \bar{C}_X - \frac{1}{2} \left[ \bar{C}_X^\dagger \bar{C}_X \rho_p + \rho_p \bar{C}_X^\dagger \bar{C}_X \right],
\] (A.18)

which is again a fairly intuitive equation as we see that the transitions can only increase the intramodule elements of \( \rho \), where the population transfer occurs, while the associated dephasing affects all elements. Separation into frequencies is trivial since the jump operators carry no time dependence, and then we can vectorize Eq. (A.18), leading to:
\[ \text{vec} \left\{ \dot{\rho}_{p|\text{trans}}^{(\omega_m)} \right\} = e^{i\omega_m t} \sum_x \left[ \delta_{p0} \left( \bar{C}_X \otimes \bar{C}_X \right) - \frac{1}{2} \left( \mathbb{1}_N \otimes \bar{C}_X^\dagger \bar{C}_X + \bar{C}_X^\dagger \bar{C}_X \otimes \mathbb{1}_N \right) \right] \text{vec} \left\{ \rho_{p}^{(\omega_m)} \right\}. \] (A.19)

Finally, the dephasing processes are the simplest to treat. Given matrices of the dephasing times in each submodule named \( T_{2,p} \), where \( T_{2,0} \) is the intramodule dephasing and \( T_{2,\pm 1} \) are the dephasings of \( \rho_{1,-1}, (T_{2,-1} = T_{2,1}^T) \), we have:

\[ \text{vec} \left\{ \dot{\rho}_{p|\text{trans}}^{(\omega_m)} \right\} = -e^{i\omega_m t} \text{vec} \left\{ \rho_{p}^{(\omega_m)} \right\} \circ \text{vec} \left\{ \rho_{p}^{(\omega_m)} \right\}, \] (A.20)

with the symbol \( \circ \) denoting the Hadamard (elementwise) product and the superscript in \( T_{2,p}^{(-1)} \) the Hadamard inverse. Now based on substitution of Eqs. (A.17), (A.19), and (A.20) into the steady-state condition of (A.15), we can organize the entire solution by the following:

\[
\begin{bmatrix}
P(H_0) + Q_1 & P(H_{-1}) & 0 \\
0 & P(H_1) & P(H_0) + Q_0 \\
0 & P(H_1) & P(H_0) + Q_{-1}
\end{bmatrix}
\times
\begin{bmatrix}
\text{vec} \left\{ \rho_{-1}^{(\omega_m)} \right\} \\
\text{vec} \left\{ \rho_0^{(\omega_m)} \right\} \\
\text{vec} \left\{ \rho_1^{(\omega_m)} \right\}
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
\] (A.21)
\[ P(H_p) = 0 \]

\[ \begin{bmatrix}
-\omega_3 & -\omega_2 & -\omega_1 & 0 & +\omega_1 & +\omega_2 & +\omega_3 \\
O_p^{-\omega_3} & O_p^{-\omega_2} & O_p^{-\omega_1} & X & X & X \\
O_p^{\omega_1} & O_p^{\omega_2} & O_p^{\omega_1} & X & X & X \\
O_p^{\omega_2} & O_p^{\omega_1} & O_p^{-\omega_1} & X & X & X \\
O_p^{\omega_1} & O_p^{\omega_2} & O_p^{\omega_1} & X & X & X \\
O_p^{\omega_3} & O_p^{\omega_2} & O_p^{\omega_1} & X & X & X \\
+\omega_1 & X & O_p^{\omega_1} & O_p^{\omega_2} & O_p^{\omega_1} & O_p^{\omega_2} \\
+\omega_2 & X & O_p^{\omega_1} & O_p^{\omega_2} & O_p^{\omega_1} & O_p^{\omega_2} \\
+\omega_3 & X & O_p^{\omega_1} & O_p^{\omega_2} & O_p^{\omega_1} & O_p^{\omega_2} \\
\end{bmatrix} \]

\[ Q_p = i \begin{bmatrix}
+\omega_3 & +\omega_2 & +\omega_1 \\
0 & -\omega_3 & -\omega_2 & -\omega_3 \\
\end{bmatrix} \otimes I_N^2 + R_p + S_p + D_p \]  

\[ O_p(\omega_m) = -\frac{i}{\hbar} \left( I_N \otimes H_p^{(\omega_m)} - H_p^{(\omega_m)}; T \otimes I_N \right) \]

\[ R_p = -i \frac{p}{\hbar} \begin{bmatrix}
-\omega_3 & -\omega_2 & -\omega_1 & 0 & +\omega_1 & +\omega_2 & +\omega_3 \\
E_{\text{mod}}^{(0)} & E_{\text{mod}}^{(-\omega_1)} & E_{\text{mod}}^{(-\omega_2)} & E_{\text{mod}}^{(-\omega_3)} & X & X & X \\
E_{\text{mod}}^{(+\omega_1)} & E_{\text{mod}}^{(0)} & E_{\text{mod}}^{(-\omega_2)} & E_{\text{mod}}^{(-\omega_3)} & X & X & X \\
E_{\text{mod}}^{(+\omega_2)} & E_{\text{mod}}^{(0)} & E_{\text{mod}}^{(-\omega_1)} & E_{\text{mod}}^{(-\omega_3)} & X & X & X \\
E_{\text{mod}}^{(+\omega_3)} & E_{\text{mod}}^{(0)} & E_{\text{mod}}^{(+\omega_1)} & E_{\text{mod}}^{(+\omega_2)} & E_{\text{mod}}^{(+\omega_3)} & E_{\text{mod}}^{(-\omega_1)} & E_{\text{mod}}^{(-\omega_2)} \\
+\omega_1 & E_{\text{mod}}^{(+\omega_1)} & X & E_{\text{mod}}^{(+\omega_2)} & E_{\text{mod}}^{(0)} & X & E_{\text{mod}}^{(-\omega_1)} \\
+\omega_2 & X & E_{\text{mod}}^{(+\omega_1)} & E_{\text{mod}}^{(+\omega_2)} & E_{\text{mod}}^{(0)} & X & E_{\text{mod}}^{(-\omega_1)} \\
+\omega_3 & X & X & E_{\text{mod}}^{(+\omega_1)} & E_{\text{mod}}^{(+\omega_2)} & E_{\text{mod}}^{(0)} & E_{\text{mod}}^{(-\omega_1)} \\
\end{bmatrix} \otimes I_N^2 \]

\[ S_p = 1 \otimes \left\{ \sum_X \left[ \delta_{p\theta} \left(C_X \otimes \bar{C}_X\right) - \frac{1}{2} \left(I_N \otimes \bar{C}_X \dagger C_X + \bar{C}_X \dagger C_X \otimes I_N \right) \right]\right\} \]

173
\[ D_p = -\mathbf{1}_7 \otimes \text{diag} \left\{ \text{vec} \left\{ T_{2,p}^{\omega(-1)} \right\} \right\}. \]

The complete steady state is then soluble after substituting a population sum condition to a single row in (A.21). The method is formulated here for a set of three frequencies where \( \omega_1 + \omega_2 = \omega_3 \), but it is straightforward to generalize to other situations.

Care must be taken to properly account for the extraction of velocities. The sum in (5.7) can be rearranged for interpretation in two different ways: if we choose \( \langle v \rangle = \sum_{ab} z_{ab} (\sum_{cd} X_{ab,cd} \rho_{cd}) \), we recover the original concept of \( \langle v \rangle = \text{Tr}(z \dot{\rho}) \), whereas if we choose instead \( \langle v \rangle = \sum_{cd} \rho_{cd} (\sum_{ab} z_{ab} X_{ab,cd}) \), it appears we have found a velocity operator \( v \) and are now using \( \langle v \rangle = \text{Tr}(v \rho) \). In this spirit, the sum scheme drawn in Fig. 5.4 can be separated into three parts based on origination from \( \rho_0 \), \( \rho_1 \), and \( \rho_{-1} \), formally:

\[ \langle v \rangle = \text{Tr}(v_0 \rho_0) + \text{Tr}(v_{-1} \rho_1) + \text{Tr}(v_1 \rho_{-1}). \]  

(A.24)

The first term can be computed directly from the pieces of (A.21) and the dipole operator, since all the pieces of the time evolution superoperator \( X \) are in place. Even the necessary frequency mixing is already organized. Using the intramodule dipole submatrix \( Z_0 \), we can evaluate the first term contributions to (A.24) at all frequencies as follows:

\[
\begin{pmatrix}
A^{(-\omega_3, -\omega_2, 0, +\omega_1, +\omega_2, +\omega_3)}
\end{pmatrix} = \left( P(H_0) + S_0 + D_0 \right) \begin{pmatrix}
\text{vec} \left\{ \rho_0^{(\omega...)} \right\}
\end{pmatrix}
\]

(A.25)

\[
\text{Tr}(v_0 \rho_0)^{(\omega...)} = \text{vec} \left\{ Z_0 \right\}^T A^{(\omega...)}.
\]

(A.26)

The second and third terms in (A.24), on the other hand, cannot be evaluated in the same approach, because the matrix equation (A.21) does not distinguish between destination modules in the evolution pointing from intramodule to intermodule elements. However, since it is seen clearly that evolution of this nature is entirely coherent (always \( P(H_{\pm1}) \)), we can be sure that the intermodule velocity operators can be constructed entirely from \( H \) and
the dipole matrix $Z$ using $v = (i/\hbar)[H, Z]$. Applying block matrix multiplication given $H$ in the form of (2.40) and a similar $Z$ having only intramodule submatrices, we obtain the off-diagonal velocity operator submatrices:

$$v_{\pm 1} = \frac{i}{\hbar} \left( [H_{\pm 1}, Z_0] \mp LH_{\pm 1} \right),$$

with $L$ as the spatial separation between modules employed in the same fashion as the energy separation in (2.40). Since $v_{\pm 1}$ carries no time dependence, evaluation of the second and third traces in (A.24) are now straightforward.
APPENDIX B

Evaluation of tunnel couplings under nonparabolicity

This Appendix shows a reliable method by which to calculate tunnel couplings under conditions of strong nonparabolicity. It is designed for use with the 3-band k.P bandstructure model in Section 2.1.4.

The problem of calculating tunnel couplings under nonparabolicity is clouded by the fact that even using the full wavefunctions as just described, we have basis functions that are not orthogonal because they come from different Hamiltonians. A reasonable approach to this is to use a two-state approximation as done in [206] and evaluate the matrix \( h = R^{-1}H \), where \( R \) is a matrix of overlap factors and \( H \) is the directly calculated Hamiltonian, between the conduction band wavefunctions only. The tunnel coupling is then evaluated as \( \Omega = \sqrt{h_{12}h_{21}} \), by comparison to a two-state tight-binding Hamiltonian. This last step that does not translate clearly to an arbitrary number of states, although for the purposes here we will keep this part of the approximation.

The novelty in this approach is that we will evaluate \( H \) and \( r \) using the full eigenstate with all three components. Applying the k.p Hamiltonian to calculate an element \( H_{12} \), we have:

\[
H_{12} = \begin{bmatrix} C_1 & L_1 & S_1 \end{bmatrix} \begin{bmatrix} E_c & \sqrt{2}\alpha & -\alpha \\ \sqrt{2}\alpha^* & E_{lh} & 0 \\ -\alpha^* & 0 & E_{so} \end{bmatrix} \begin{bmatrix} C_2 \\ L_2 \\ S_2 \end{bmatrix}.
\]

Evaluating this equation gives:
Figure B.1: Accuracy of the method for calculation of tunnel couplings under strong non-parabolicity.

\[
H_{12} = \langle C_1|E_c|C_2 \rangle + \langle L_1|E_{lh}|L_2 \rangle + \langle S_1|E_{so}|S_2 \rangle \\
+ \hbar \langle C_1 \rangle \frac{\partial}{\partial z} \left[ \sqrt{\frac{E_p}{2m_0}} (\sqrt{2}|L_2 \rangle - |S_2 \rangle) \right] - \hbar \left[ \sqrt{2}\langle L_1 \rangle - \langle S_1 \rangle \right] \sqrt{\frac{E_p}{2m_0}} \frac{\partial}{\partial z}|C_2 \rangle.
\]

The terms in the top row are clearly potential energy, which mix like-components, and the terms in the bottom are kinetic energy, which is determined purely by mixing of different components. Taken directly, the above equation gives a straightforward way to calculate \( H \).

As an example we consider a coupled pair of wells in In\textsubscript{0.53}Ga\textsubscript{0.47}As/In\textsubscript{0.52}Al\textsubscript{0.48}As lattice matched to InP, with the well width fixed at 4 nm and sweeping the barrier. We compare the calculated ground state tunnel coupling with the actual anticrossing gap (divided by two) obtained by solving the actual two-well system in Figure B.1: this method yields a coupling energy which is larger than the anticrossing by less than 0.2% for barrier widths in the 3-5 nm range and staying still below 2% for extremely thin barriers; this is a significant improvement over the method using only the conduction band wavefunction, which underpredicts the tunnel couplings by around 20% [206].

177
APPENDIX C

Future experiment: nanopillar intraband photoconductor

An immediate milestone in the development of a THz QDCL will be the observation of intraband optical transitions in nanopillars. This Appendix proposes a device for such an experiment, named a nanopillar intraband photoconductor (NIP). Device concept, electromagnetic simulations, and fabrication details are given.

C.1 Design concept

Figure C.1 shows the NIP device concept, the intent of which is to act as a photoconductor at mid-IR or THz wavelengths due to intraband optical phenomena. Large-area arrays of nanopillar axial heterostructures are grown on a doped substrate on a lattice defined by electron-beam lithography (selective-area epitaxy), and are covered by a metal contact with periodic holes to act as both a top electrical contact and an optical coupler for normally-incident radiation from the top. Under current capabilities, the array sizes can practically reach up the \((100s \text{ of } \mu \text{m})^2\), and at pitches down to 400 nm, so that each device can have potentially hundreds of thousands of pillars. The advantage of using such a large sample size of pillars is that the signal strength is increased by the number of pillars \(N\), and the signal-to-noise ratio is increased by approximately \(\sqrt{N}\) (ignoring systematic fluctuations across the device area). This would hopefully boost the photoconductivity signal quality so as to make photoconductance spectroscopy possible using FTIR.

The pillar arrays must be planarized for the top contact to sit, which can be done for example using the polymer Benzocyclobutene (BCB). BCB can be spun on at a thickness
greater than the pillar heights and then etched back to reveal the tips, after which the 
top contact is patterned and deposited. The hole size and spacing in the top contact can 
be selected so that specific surface plasmon resonances exist at chosen frequencies on the 
bottom side, which are the mechanism of light coupling to the NP heterostructure array. The 
target periodicities are roughly the wavelength of the target frequency in the BCB/nanopillar 
effective medium, which might extend down to roughly a few $\mu$m for the mid-IR. Therefore, 
the pillar spacing is far below the hole spacing, so that many uncontacted pillars will sit 
within each hole.

Coupling light to surface plasmons using the grating has multiple advantages. One ad-

tantage is that top-side illumination is made possible, rather than back-side illumination 
through the substrate which would not be possible using a doped substrate. Additionally, 
the grating automatically provides a high degree of polarization rotation: surface plasmons 
inherently have a high degree of vertical polarization due to the boundary conditions on the 
metal-dielectric interface, and this vertical polarization would presumably be necessary to 
excite axial transitions in axial NP heterostructures. In certain experiments using quantum 
wells this constraint has required illumination through an angled cleaved edge of the sub-
strate, which would again not be possible if the substrate were doped. In fact, even with an 
undoped substrate, a backside illumination approach would still need some type of grating
scheme to produce vertical polarization.

Similar strategies have been used to observe photoconductivity in other low-dimensional materials. In [207], an interband nanopillar detector was made where the grating periodicity is defined by the pillar spacing so as to achieve surface plasmon resonances in the near-IR. In [208], a surface plasmon grating was used to enhance the responsivity of a quantum-dot infrared photodetector composed of layers of self-assembled quantum dots. Similarly in [209], such high field enhancement was achieved that strong coupling was observed.

In our application, the NIP concept has the advantage of versatility, which will hopefully allow it to be repeated as an intraband optical characterization technique as growth capabilities develop. This includes all types of axial heterostructures (for example single-barrier, double-barrier, superlattice, etc.), and frequencies throughout the mid-IR and THz.

C.2 Electromagnetic simulations

Electromagnetic simulations were performed using the software package HFSS to validate the effectiveness of the surface plasmon grating. The natural figure of merit is the amount of useable optical energy in comparison to the incident plane-wave energy, where “useable” is defined as the integrated energy coming from z-directed polarization under only the metal-contacted areas. Specifically, this is:

\[
F = \frac{\frac{1}{2} \varepsilon_{zz} \int_{\text{metal}} |E_z|^2 dV}{ILLnT/c}.
\]  

(C.1)

The numerator is the total z-polarized electromagnetic energy housed under the metal, and the denominator is the total electromagnetic energy of the incident plane wave over the length \(L\) of the pillars, corrected by the transmission coefficient \(T\) between air and the active NP/BCB regions. \(I\) is the incident intensity.

The active NP/BCB region is modeled as an anisotropic effective medium using Maxwell-Garnett effective medium theory [210, 211], resulting in a different permittivity in the in-plane directions \((\varepsilon_{xx} = \varepsilon_{yy})\) than in the vertical direction \((\varepsilon_{zz})\). For InAs pillars embedded
in BCB, we obtain $\epsilon_{xx} = \epsilon_{yy} = 3.53$ and $\epsilon_{zz} = 2.92$. The refractive index $n$ is taken from the in-plane components, and $T$ is calculated from $n$.

Some simulated results are shown in Figure C.2, for examples aimed at the mid-IR (left) and THz (right). $L$ is taken as 1.5 $\mu$m. For the mid-infrared, specific plasmon resonances are labeled which occur when the wavevector in the medium is equal to multiples of the contact lattice vectors; these are labeled by $\text{SP}_{mn}$, where $m, n$ are the lattice wavevector indices. These resonances are broadened sufficiently by radiative losses that a spectrum of coupling can be engineered between approximately 6-11 $\mu$m covered by the semi-merging SP$_{10}$ and SP$_{11}$ resonances. This range is sufficient for a photoconductance spectrum, although care should be taken to deconvolve the grating spectral features from the results. The bottom middle plot shows the intensity distribution of z-polarized radiation in the surface at the SP$_{10}$ resonance, and over a planar surface which is cut diagonal to the incident polarization. Field localization occurs near the contact edges, and the decay length appears to be around 1 $\mu$m, which is convenient for realistic pillar lengths.

At THz wavelengths, the surface mode attains a more waveguide-like character, due to
The longer wavelength and the metallic nature of the substrate at lower frequency. These modes are TM-polarized due to the short waveguide cross section, and so have a high degree of vertical polarization. As was the case in the mid-IR, the resonances are broadened by radiative losses. The simulations suggests that the scheme can allow for broadband coupling over the entire range of approximately 2-7 THz, although it will still be necessary to deconvolve the resonant nature of the grating response from that of any intraband photoconductive features.

C.3 Fabrication details

Figure C.3 shows an overview of a fabrication process for a NIP, from substrate to device. Steps 1-7 follow the usual technique of selective-area epitaxy, after which planarization is performed over steps 8-9, and the grating contact is formed over steps 10-13. Electron-beam
lithography (e-beam) is chosen as usual for the growth pattern, and in this process the
grating contact is also patterned using e-beam. This is necessary for example in the mid-
infrared to produce the \( \mu \text{m} \)-scale features, although in fabrication of a larger THz grating
optical lithography should be acceptable to reduce cost. Nanoimprint lithography could also
potentially be used to pattern both mid-IR and THz gratings.
References


[179] B.R. Nag, “Interface roughness scattering limited mobility in AlAs/GaAs, Al0.3Ga0.7As/GaAs and Ga0.5In0.5P/GaAs quantum wells,” Semiconductor Science and Technology 19, 162 (2003).


