Characterization of mid-infrared quantum cascade lasers

by

David Patrick Burghoff

B.S., University of Illinois (2007)

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

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Author . . . .
Department of Electrical Engineering and Computer Science
September 4, 2009

Certified by . . . .
Qing Hu
Professor
Thesis Supervisor

Accepted by . . . .
Terry P. Orlando
Chairman, Department Committee on Graduate Theses

Characterization of mid-infrared quantum cascade lasers

by

David Patrick Burghoff

Submitted to the Department of Electrical Engineering and Computer Science
on September 4, 2009, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

Quantum cascade lasers provide some of the highest output powers available for light in the mid-infrared range (from 3 to 8 μm). As many of their applications require portability, designs that have a high wall-plug efficiency are essential, and were designed and grown by others to achieve this goal. However, because a large fraction of these devices did not operate at all, very few of the standard laser measurements could be performed to determine their properties. Therefore, measurements needed to be performed that could non-destructively probe the behavior of QCLs while still providing useful information.

This thesis explores these types of measurements, all of which fall into the category of device spectroscopy. Through polarization-dependent transmission and photovoltaic spectroscopy, a large portion of the quantum mechanical band structure could be determined, along with many of the parameters characterizing crystal growth quality. In addition, high-resolution transmission spectroscopy was used to find the properties of the QCL waveguide. In order to find the correspondence between theory and experiment, band structure simulations were performed using a three-band $k \cdot p$ model, and two-dimensional electromagnetic simulations were performed to describe the laser's optical properties. These simulations were found to be in relatively good agreement with the device measurements, and any discrepancies were found to be consistent with problems in the growth and fabrication.

Thesis Supervisor: Qing Hu
Title: Professor

Acknowledgments

First and foremost, I'd like to thank my research advisor, Professor Qing Hu, for the financial support and guidance he's given me on this project. I'd also like to acknowledge the current and former members of Prof. Hu's research group, especially Allen Hsu and Alan Lee, but also Wilt Kao, Sushil Kumar, Qi Qin, Ivan Chan, and visiting researcher Prof. Dayan Ban. I thank our collaborators in Dr. George Turner's group at MIT Lincoln Laboratory for growing and fabricating the devices that were tested, and also Dr. Dmitry Revin at the University of Sheffield for his helpful correspondence in setting up the transmission experiment. Lastly, I'd like to thank my mother, father, brother, and sisters for offering me unwavering moral support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since their invention in 1994, quantum cascade lasers (QCLs) have proven to be an extremely important source of long-wavelength light. [1] Their versatility is unmatched amongst semiconductor lasers, having been demonstrated for wavelengths as short as those in the mid-infrared, and as long as those in the sub-millimeter regime. In other words, QCLs fill the sizable gap that lies between conventional bipolar semiconductor lasers and high-frequency electronics. A QCL is created by growing alternating layers of different materials on a semiconductor substrate, including lattice-matched AlGaAs/GaAs or strained InGaAs/InAlAs. Because the layers are only a few monolayers thick, quantum-size effects dominate, and an artificial band structure is formed, permitting the designer to tailor the design wavelength simply by changing the growth thicknesses. This is a stark contrast to the conventional bipolar laser, in which the device's properties are largely fixed by the material's band gap. As a result, QCLs do not have an intrinsic upper bound on the wavelengths they can generate, and are tremendously versatile.

Specifically, QCLs operating in the mid-infrared (3-8 microns) regime have many novel applications, ranging from chemical sensing via spectroscopy to IR countermeasures. [2] One constraint that has limited the development of these applications is the lack of portability associated with the high power requirements and low wall-plug efficiencies. To that end, designs have been proposed by our group for efficient "injectorless" designs as part of the DARPA Efficient Mid-Infrared Lasers (EMIL) pro-

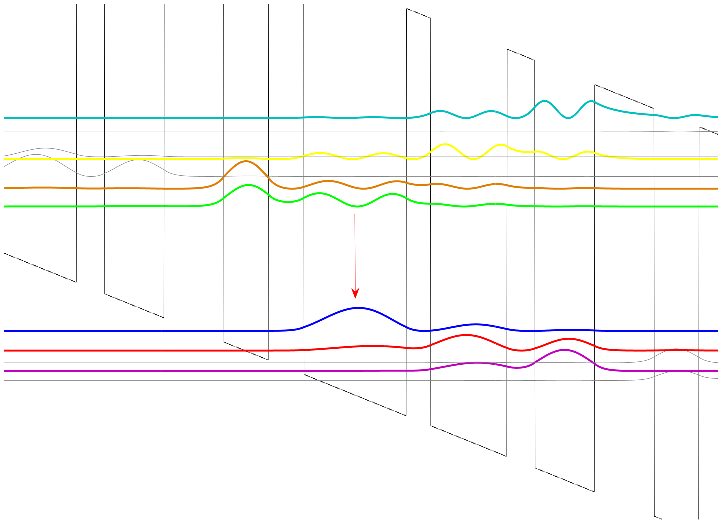


Figure 1-1: Basic schematic of a QCL under bias.

gram. [3] These structures were grown by metal-organic chemical vapor deposition (MOCVD) and fabricated at MIT Lincoln Laboratory, along with some previously published test structures. However, certain problems arose that suggested that something went awry in the growth process. Certain devices were grown to be identical to already published designs, yet they did not lase. In addition, almost none of the novel designs actually operated; those that did operate did so at a wavelength significantly different from the design wavelength. This work will attempt to diagnose some of these issues through the use of varying types of spectroscopy. Spectroscopy is an extremely powerful tool that allows one to extract many of the parameters relevant to QCL design, including the band structure, doping levels, effective mass and non-parabolicity factors, conduction band offset, and oscillator strengths. Moreover, it permits the measurement of certain parameters that may otherwise be difficult to directly measure, including active region temperature and waveguide losses.

There are several measurement techniques that can provide information about a laser's performance. The most straightforward include current-voltage (IV), light-current (LI), and spectral measurements. While these techniques are certainly valuable evaluation tools, they are ultimately functions of many parameters and contain few specifics about any one in particular. Moreover, they typically require working lasers in order to give any useful information. For devices which do not lase, electroluminescence (EL) measurements can provide a little more information about a device's band structure, but they are ultimately limited by the fact that only states undergoing a significant population inversion will emit light. Ultimately, EL will provide information similar to spectral measurements on a lasing device, but only information pertaining to a single pair of states.

Another type of measurement commonly performed are the related Hakki-Paoli techniques and the internal Fourier-Hakki-Paoli techniques. [4] [5] In each case, the amplified spontaneous emission spectrum of a biased device is measured by a spectrometer, and the interference of multiple bounces in the cavity is used to measure the gain spectrum. In the former case, the size of interference fringes on the spectrum is used to determine the net gain in the active region as a function of wavelength and bias. In the latter case, interference results in several spatially-separated copies of the spontaneous emission spectrum on the interferogram (the Fourier transform of the spectrum). These copies are divided and inverse-transformed in order to recover the net gain. While the second method has better noise performance and is generally preferable to the first [6], both are somewhat limited by the fact that they require high spontaneous emission signals, thereby requiring devices be operated in continuous-wave (CW) mode. For devices with poor heat-sinking, this may not be possible without destroying the device. In addition, both methods also require that the devices operate in the amplified-spontaneous-emission regime, and so they must be operating at a bias close to the laser threshold. Not only is this impossible for devices that do not lase in the first place, but it also strictly limits the bias range in which the device's gain spectrum can be measured.

Photodetection provides another tool through which a device's parameters can be measured. [7] By biasing a device at different current levels, using the device as a photodetector, and measuring the spectrum of a high-bandwidth light source that has passed through an interferometer with the detecting device, one can easily extract the responsivity of the detecting device as a function of wavelength. From this, one can determine the energies of transitions that the device's band structure can support. The main complication associated with this is twofold. Firstly, because the coupling efficiency of light into the device from an incoherent source is largely unknown, responsivity values can only be found in relative (and not absolute) terms. Thus, no absolute quantitative measurements are possible. Secondly, the responsivity is dependent not only upon the oscillator strengths of individual levels, but also upon the ability of photoexcited electrons to transit from the active region into the contacts. In general, this is a complicated function of the bias and the internal dynamics of the device; modeling requires the sophisticated simulation techniques like Monte-Carlo. Since there are so many parameters that are required to describe the transit ability, extraction of any one parameter (such as conduction band offset or effective mass) is difficult.

An extremely powerful and versatile tool for device characterization which lacks many of the limitations of the previous methods is the demonstrated usage of broadband transmission spectroscopy. [8] Like the photodetectivity measurement set-up, light from an incoherent blackbody source is coupled into one facet of a biased device. It is then collected from the other side, passed through a polarizer, and the resulting light is measured with a spectrometer. Owing to the well-known polarization selection rule in QCLs, only TM-polarized light will be affected by the device's active region, whilst TE-polarized light will pass through uninhibited. Since waveguide losses are typically similar for each polarization, dividing and scaling the resulting spectra results in the absolute absorption of the active region alone (sans waveguide losses). The blackbody source can be modulated to remove the effects of emission from the active region itself, and the method does not require any population inversion, so the method is applicable to a device under a wide range of biases, from zero all the way up past lasing threshold. By varying the applied bias and measuring, this method permits the identification of energy states with low population densities and low transit ability.

This thesis describes the simulations and measurements that were performed to characterize the important properties of different QCLs. Chapter 2 details the quantum mechanics that underlies the subband structure of a QCL and its interaction with light. Chapter 3 describes the numerical simulations that were performed in order to complete understanding of the band structure, as well as the important electromagnetic simulations of the laser waveguides. Chapter 4 portrays the experimental setups used for the various measurement types. Finally, Chapter 5 includes the results of the measurements and simulations, as well as the analysis of their implications on laser performance.

Chapter 2

QCL Theory

In order to properly characterize the observed behavior of QCLs, a robust theoretical description of the underlying quantum mechanics is essential. To that end, a formalism will be used first developed by Kane [9] to describe III-V materials and later expanded upon by Bastard [10] and Sirtori [11] to describe heterostructures.

2.1 Band structure of III-V materials

Though there are many ways to calculate the band structure of solids, the 8×8 k·p method of Kane is one of the most powerful, because it simplifies greatly in the presence of heterostructures, unlike more sophisticated descriptions such as those involving pseudopotentials or density-functional theory. For bulk solids, the electron eigenstates must satisfy Schrödinger's equation [10]:

$$\mathcal{H}_0\vec{\Psi}(\mathbf{r}) \equiv \left[\frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r}) + \frac{\hbar}{4m_0^2c^2}(\boldsymbol{\sigma} \times \nabla V) \cdot \mathbf{p}\right] \vec{\Psi}(\mathbf{r}) = E\vec{\Psi}(\mathbf{r}), \tag{2.1}$$

where the first term represents kinetic energy, the second represents potential energy, and the third represents spin-orbit coupling. Note that the wavefunction has been denoted with a vector arrow to emphasize that spin has not been neglected, and the wavefunction has been taken to have two components. In a crystal, $V(\mathbf{r})$ is a periodic function of space, so Bloch's theorem applies, and the eigenstates are known to be of the form

$$\begin{aligned} \vec{\Psi}_{n\mathbf{k}}(\mathbf{r}) &= \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{r}} \vec{u}_{n\mathbf{k}}(\mathbf{r}) \\ \mathcal{H}_0 \vec{\Psi}_{n\mathbf{k}}(\mathbf{r}) &= E_n(\mathbf{k}) \vec{\Psi}_{n\mathbf{k}}(\mathbf{r}), \end{aligned} \tag{2.2}$$

where n is the band index, $\hbar\mathbf{k}$ is the crystal momentum, N is the number of unit cells, and $\vec{u}_{n\mathbf{k}}(\mathbf{r}) = \vec{u}_{n\mathbf{k}}(\mathbf{r}+\mathbf{R})$ is a Bloch function with the same periodicity as the lattice. Noting that $\mathbf{p}\left(e^{i\mathbf{k}\cdot\mathbf{r}}\vec{u}_{n\mathbf{k}}(\mathbf{r})\right) = e^{i\mathbf{k}\cdot\mathbf{r}}(\mathbf{p}+\hbar\mathbf{k})\vec{u}_{n\mathbf{k}}(\mathbf{r})$, one can use (2.2) in (2.1) to obtain

$$\mathcal{H}_{\mathbf{k}} \vec{u}_{n\mathbf{k}} \equiv \left[\mathcal{H}_0 + \frac{\hbar^2 \mathbf{k}^2}{2m_0} + \frac{\hbar}{m_0}\mathbf{k}\cdot\mathbf{p} + \frac{\hbar}{4m_0^2 c^2}(\boldsymbol{\sigma}\times\nabla V)\cdot\hbar\mathbf{k}\right]\vec{u}_{n\mathbf{k}} = E_n(\mathbf{k})\vec{u}_{n\mathbf{k}} \tag{2.3}$$

In other words, the complicated problem of solving Schrödinger's equation over all unit cells has been reduced to one of solving it over a single cell (under periodic boundary conditions). The downside to this is that for each value of $\mathbf{k}$, a different Hamiltonian $\mathcal{H}_{\mathbf{k}}$ needs to be solved. Fortunately, each of the $\mathcal{H}_{\mathbf{k}}$ operators are Hermitian, meaning that if one can find all of the eigenstates at a single value of $\mathbf{k}$ (say $\mathbf{k} = 0$), one can in principle expand the eigenstates of *every* $\mathcal{H}_{\mathbf{k}}$ using the states at a single $\mathbf{k}$ as a basis.

Kane's contribution was to realize that in the case of III-V semiconductors, the Bloch wavefunctions near the Γ point are primarily comprised of eight well-characterized bands: the six p-like valence bands and the two s-like conduction bands. In that case, it is generally sufficient to truncate what would otherwise be an infinite expansion into a finite one:

$$\vec{u}_{n\mathbf{k}}(\mathbf{r}) = \sum_{m=1}^{8} d_{nm}(\mathbf{k})\vec{u}_{n0}(\mathbf{r}) \tag{2.4}$$

(where the $d_{nm}(\mathbf{k})$'s are simply expansion coefficients). The analytical forms of the $\vec{u}_{n\mathbf{0}}$'s can be found by solving (2.3) for $\mathbf{k} = 0$, and have been included in Table 2.1. In the table, the energies E_C, E_g, and Δ are respectively called the conduction band offset, energy gap, and split-off energy, and the states labeled as S, X, Y, and Z are states with the same symmetry as atomic s-states and p-states.

To find the dispersion relation as a function of $\mathbf{k}$, one must use (2.4) to find

Table 2.1: Γ-point Bloch functions

$\vec{u}_{n\mathbf{0}}$	$\lvert J, m_J\rangle$	Wavefunction:	Energy:	Name:
$\vec{u}_{1\mathbf{0}}$	$\lvert 1/2, 1/2\rangle$	$i\lvert S\uparrow\rangle$	E_C	Conduction ↑
$\vec{u}_{2\mathbf{0}}$	$\lvert 3/2, 1/2\rangle$	$-\sqrt{\frac{2}{3}}\lvert Z\uparrow\rangle + \frac{1}{\sqrt{6}}\lvert (X+iY)\downarrow\rangle$	$E_C - E_g$	Light-hole ↑
$\vec{u}_{3\mathbf{0}}$	$\lvert 3/2, 3/2\rangle$	$\frac{1}{\sqrt{2}}\lvert (X+iY)\uparrow\rangle$	$E_C - E_g$	Heavy-hole ↑
$\vec{u}_{4\mathbf{0}}$	$\lvert 1/2, 1/2\rangle$	$\frac{1}{\sqrt{3}}\lvert (X+iY)\downarrow\rangle + \frac{1}{\sqrt{3}}\lvert Z\uparrow\rangle$	$E_C - E_g - \Delta$	Split-off ↑
$\vec{u}_{5\mathbf{0}}$	$\lvert 1/2, -1/2\rangle$	$i\lvert S\downarrow\rangle$	E_C	Conduction ↓
$\vec{u}_{6\mathbf{0}}$	$\lvert 3/2, -1/2\rangle$	$-\sqrt{\frac{2}{3}}\lvert Z\downarrow\rangle - \frac{1}{\sqrt{6}}\lvert (X-iY)\uparrow\rangle$	$E_C - E_g$	Light-hole ↓
$\vec{u}_{7\mathbf{0}}$	$\lvert 3/2, -3/2\rangle$	$\frac{1}{\sqrt{2}}\lvert (X-iY)\downarrow\rangle$	$E_C - E_g$	Heavy-hole ↓
$\vec{u}_{8\mathbf{0}}$	$\lvert 1/2, -1/2\rangle$	$-\frac{1}{\sqrt{3}}\lvert (X-iY)\uparrow\rangle + \frac{1}{\sqrt{3}}\lvert Z\downarrow\rangle$	$E_C - E_g - \Delta$	Split-off ↓

the eigenstates of $\mathcal{H}_{\mathbf{k}}$. If the $\vec{u}_{n\mathbf{0}}$'s have been normalized over the unit cell (i.e., $\int_{V_0} \vec{u}^*_{n\mathbf{0}} \cdot \vec{u}_{n\mathbf{0}}\, d^3\mathbf{r} = 1$), then the matrix representation of $\mathcal{H}_{\mathbf{k}}$, denoted by the 8×8 matrix $\overline{\overline{\mathcal{H}}}(\mathbf{k})$, is simply given by its elements

$$\mathcal{H}_{mn}(\mathbf{k}) \equiv \langle \vec{u}_{m\mathbf{0}}\rvert \mathcal{H}_{\mathbf{k}} \lvert \vec{u}_{n\mathbf{0}}\rangle \equiv \int_{V_0} \vec{u}^*_{m\mathbf{0}} \cdot \mathcal{H}_{\mathbf{k}} \vec{u}_{n\mathbf{0}}\, d^3\mathbf{r} \tag{2.5}$$

The actual calculation of these elements is tedious, but is greatly simplified by the choice of basis states. In fact, thanks to the symmetry of the states, many of the terms vanish. The calculated value of $\overline{\overline{\mathcal{H}}}(\mathbf{k})$ is given by

$$\begin{pmatrix} E_C+\frac{\hbar^2\mathbf{k}^2}{2m_0} & -\sqrt{\frac{2}{3}}P\hbar k_z & P\hbar k_+ & \frac{1}{\sqrt{3}}P\hbar k_z & 0 & -\frac{1}{\sqrt{3}}P\hbar k_- & 0 & -\sqrt{\frac{2}{3}}P\hbar k_- \\ -\sqrt{\frac{2}{3}}P\hbar k_z & E_{LH}+\frac{\hbar^2\mathbf{k}^2}{2m_0} & 0 & 0 & \frac{1}{\sqrt{3}}P\hbar k_- & 0 & 0 & 0 \\ P\hbar k_- & 0 & E_{HH}+\frac{\hbar^2\mathbf{k}^2}{2m_0} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{\sqrt{3}}P\hbar k_z & 0 & 0 & E_{SO}+\frac{\hbar^2\mathbf{k}^2}{2m_0} & \sqrt{\frac{2}{3}}P\hbar k_- & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}}P\hbar k_+ & 0 & \sqrt{\frac{2}{3}}P\hbar k_+ & E_C+\frac{\hbar^2\mathbf{k}^2}{2m_0} & -\sqrt{\frac{2}{3}}P\hbar k_z & P\hbar k_- & \frac{1}{\sqrt{3}}P\hbar k_z \\ -\frac{1}{\sqrt{3}}P\hbar k_+ & 0 & 0 & 0 & -\sqrt{\frac{2}{3}}P\hbar k_z & E_{LH}+\frac{\hbar^2\mathbf{k}^2}{2m_0} & 0 & 0 \\ 0 & 0 & 0 & 0 & P\hbar k_+ & 0 & E_{HH}+\frac{\hbar^2\mathbf{k}^2}{2m_0} & 0 \\ -\sqrt{\frac{2}{3}}P\hbar k_+ & 0 & 0 & 0 & \frac{1}{\sqrt{3}}P\hbar k_z & 0 & 0 & E_{SO}+\frac{\hbar^2\mathbf{k}^2}{2m_0} \end{pmatrix}, \tag{2.6}$$

where $E_{LH} \equiv E_{HH} \equiv E_C - E_g$ are the light- and heavy-hole energies, $E_{SO} \equiv E_C - E_g - \Delta$ is the split-off energy, $P \equiv -i/m_0 \langle S\rvert p_x \lvert X\rangle$ is the interband momentum matrix element, and $k_\pm \equiv \frac{1}{\sqrt{2}}(k_x \pm ik_y)$. [10] All things considered, the above matrix is quite simple: it has only four material-dependent parameters (E_C,E_g, Δ, and P), all of which are tabulated and widely available. In addition, essentially all of the geometry of the system has vanished, eliminating the need to model a complicated atomic potential. Once the four parameters are known, finding $E_n(\mathbf{k})$ becomes a straightforward matrix eigenvalue problem, one that cannot be solved analytically but can that be handled numerically quite easily.

2.2 Envelope functions and heterostructures

Though $\overline{\overline{\mathcal{H}}}(\mathbf{k})$ is useful for describing the properties of bulk material, it lacks descriptive power in the face of any system where the crystal periodicity is broken. Any heterostructure device will fall into this category, so to describe these, one needs to consider wavepackets consisting of a superposition of many different Bloch states, $\sum_{n,\mathbf{k}} c_n(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\vec{u}_{n\mathbf{k}}(\mathbf{r})$. However, if Kane's fundamental assumption is valid, then the $\vec{u}_{n\mathbf{k}}$ states can themselves can be expressed as a superposition of the $\vec{u}_{n0}$ states. In other words, an arbitrary wavefunction can be taken to be of the form

$$\vec{\Psi} = \sum_{n=1}^{8}\sum_{\mathbf{k}} c_n(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\vec{u}_{n0}(\mathbf{r}) = \sum_{n=1}^{8}\left(\sum_{\mathbf{k}} c_n(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\right)\vec{u}_{n0}(\mathbf{r}). \tag{2.7}$$

Observing that the quantity in parentheses varies slowly compared to the scale of a unit cell for values of $\mathbf{k}$ near the Γ-point, one can rewrite this as

$$\vec{\Psi} = \sum_{n=1}^{8} F_n(\mathbf{r})\vec{u}_{n0}(\mathbf{r}), \tag{2.8}$$

where each $F_n(\mathbf{r})$ is just a slowly-varying envelope function. Though (2.8) is often referred to as the envelope approximation [3], the only approximation that has been made is Kane's, and in principle, if one were to extend this definition to include an infinite number of bands, it would become exact.

Now, suppose that a slowly-varying potential $V_{ext}(\mathbf{r})$ is applied to bulk material.

In this case, Schrödinger's Equation can be rewritten as

$$\begin{aligned}
0 &= [\mathcal{H}_0 + V_{ext} - E] \sum_{n,\mathbf{k}} c_n(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \vec{u}_{n0}(\mathbf{r}) \\
&= \sum_{n,\mathbf{k}} c_n(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} [\mathcal{H}_{\mathbf{k}} + V_{ext} - E] \vec{u}_{n0}(\mathbf{r}). \qquad (2.9)
\end{aligned}$$

Evaluating both sides of the above equation at $\mathbf{r} = \mathbf{r}'$, multiplying both sides by the pseudo-completeness operator $\sum_{\mathbf{k}'} \frac{1}{V} e^{-i\mathbf{k}'\cdot\mathbf{r}} \int d^3\mathbf{r}' e^{-i\mathbf{k}'\cdot\mathbf{r}'} \vec{u}^*_{m0}(\mathbf{r}')\cdot$ and performing a considerable amount of algebra, it can be shown that (2.9) reduces to the following Schrödinger-like equation:

$$0 = \sum_n (\mathcal{H}_{mn}(-i\nabla) + (V_{ext} - E)\delta_{mn}) F_n(\mathbf{r}). \qquad (2.10)$$

Since the above statement holds for all m, it can also be rewritten compactly in matrix form as

$$\left[\overline{\overline{\mathcal{H}}}(-i\nabla) + V_{ext}(\mathbf{r})\overline{\overline{I}}\right] \vec{F}(\mathbf{r}) = E\vec{F}(\mathbf{r}), \qquad (2.11)$$

where $\vec{F}(\mathbf{r})$ is the eight-component column vector containing each envelope function and $\overline{\overline{I}}$ is the identity matrix. Unsurprisingly, the equation for the envelope functions has reduced to an eigenvalue problem, albeit one with differential operators that is in general quite difficult to solve.

Technically speaking, equation (2.11) only applies to perturbations applied to bulk materials. If one wants to model heterostructures, the above analysis needs to be modified to include those effects. Suppose that a system is comprised of two materials, A and B. Defining the modified Bloch functions as $\vec{u}^{A/B}_{n0}(\mathbf{r}) \equiv \vec{u}^{A}_{n0}(\mathbf{r})U_A(\mathbf{r}) + \vec{u}^{B}_{n0}(\mathbf{r})U_B(\mathbf{r})$, where $U_A(\mathbf{r})$ and $U_B(\mathbf{r})$ are unit step functions indicating the presence of an A material or a B material, and $\vec{u}^{A}_{n0}$ and $\vec{u}^{B}_{n0}$ are the respective Bloch functions, one can then take the wavefunctions to be of the form

$$\vec{\Psi} = \sum_{n=1}^{8} F_n(\mathbf{r}) \vec{u}^{A/B}_{n0}(\mathbf{r}). \qquad (2.12)$$

Assuming that each individual layer retains its structure, the previous analysis will remain relatively unchanged, and one can show that (2.11) changes to become

$$\left[\overline{\overline{\mathcal{H}}}_A(-i\nabla)U_A(\mathbf{r}) + \overline{\overline{\mathcal{H}}}_B(-i\nabla)U_B(\mathbf{r}) + V_{ext}(\mathbf{r})\overline{\overline{I}}\right]\vec{F}(\mathbf{r}) = E\vec{F}(\mathbf{r}). \tag{2.13}$$

The heterostructure merely creates position-dependent material parameters, and thus as long as proper continuity-of-wavefunction and continuity-of-probability-current boundary conditions are used at the interfaces, the matrix can be diagonalized in a manner similar to the one-material case.

Lastly, it will be useful to determine how the quantum mechanical inner product manifests for wavefunctions expressed in terms of envelope functions. Using (2.8) in the definition of the inner product for wavefunctions with spin, and assuming that the Bloch functions have been normalized so that $\int_{V_0} \vec{u}^*_{n\mathbf{0}} \cdot \vec{u}_{n\mathbf{0}}\, d^3\mathbf{r} = V_0$ (where V_0 is the volume of a unit cell),

$$\begin{aligned}
\left\langle \vec{\Psi}_1 \middle| \vec{\Psi}_2 \right\rangle &= \int \left(\sum_{n=1}^{8} F_n^{(1)}(\mathbf{r})\vec{u}_{n0}(\mathbf{r})\right)^* \cdot \left(\sum_{n=1}^{8} F_n^{(2)}(\mathbf{r})\vec{u}_{n0}(\mathbf{r})\right) d^3\mathbf{r} \\
&= \sum_{n,m} \sum_{jth\,unit\,cell} \int_j F_n^{(1)^*}(\mathbf{r})F_m^{(2)}(\mathbf{r})\vec{u}^*_{n0}(\mathbf{r}) \cdot \vec{u}_{m0}(\mathbf{r})d^3\mathbf{r} \\
&\approx \sum_{n,m} \sum_{jth\,unit\,cell} F_n^{(1)^*}(\mathbf{r}_j)F_m^{(2)}(\mathbf{r}_j) \int_j \vec{u}^*_{n0}(\mathbf{r}) \cdot \vec{u}_{m0}(\mathbf{r})d^3\mathbf{r} \\
&= \sum_{n,m} \sum_{jth\,unit\,cell} F_n^{(1)^*}(\mathbf{r}_j)F_m^{(2)}(\mathbf{r}_j)V_0\delta_{nm} \\
&\approx \sum_n \int F_n^{(1)^*}(\mathbf{r})F_n^{(2)}(\mathbf{r})d^3\mathbf{r} = \int \vec{F}_1^* \cdot \vec{F}_2\, d^3\mathbf{r}.
\end{aligned} \tag{2.14}$$

This process, in which a slowly-varying function is pulled out of an integral and the resulting sum is approximated by an integral, is a common method of dealing with envelope functions. (In fact, it was implicitly used in the derivation of (2.10).) Perhaps not surprisingly, all of the microscopic dependence of the wavefunction has vanished from the product. In addition, note that the strange choice of normalization for the Bloch functions allows one to normalize the total function by collectively normalizing its component envelopes. This greatly simplifies the calculation of matrix elements.

2.3 One-dimensional heterostructures

Though the Schrödinger-like equation presented in (2.13) has been greatly simplified from its original version, it nevertheless remains quite daunting. Fortunately, quantum cascade lasers have very little variation beyond that in the growth direction (denoted by $\hat{z}$), in which case the effective Hamiltonian is translation-invariant and the envelope functions can be taken to be of the form

$$\vec{F}(\mathbf{r}) = \frac{1}{\sqrt{A}}\, e^{i(k_x x + k_y y)} \vec{F}_{k_x,k_y}(z). \tag{2.15}$$

In most cases, the main wavefunctions of interest will be those for which $k_x = k_y = 0$. When this is the case, any terms in $\overline{\overline{\overline{\mathcal{H}}}}(-i\nabla)$ corresponding to transverse variation will drop out, leaving an $\overline{\overline{\overline{\mathcal{H}}}}(k_z = -i\frac{\partial}{\partial z})$ given by

$$\begin{pmatrix}
E_C+\frac{\hbar^2k_z^2}{2m_0} & -\sqrt{\frac{2}{3}}P\hbar k_z & 0 & \frac{1}{\sqrt{3}}P\hbar k_z & 0 & 0 & 0 & 0 \\
-\sqrt{\frac{2}{3}}P\hbar k_z & E_{LH}+\frac{\hbar^2k_z^2}{2m_0} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & E_{HH}+\frac{\hbar^2k_z^2}{2m_0} & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{\sqrt{3}}P\hbar k_z & 0 & 0 & E_{SO}+\frac{\hbar^2k_z^2}{2m_0} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & E_C+\frac{\hbar^2k_z^2}{2m_0} & -\sqrt{\frac{2}{3}}P\hbar k_z & 0 & \frac{1}{\sqrt{3}}P\hbar k_z \\
0 & 0 & 0 & 0 & -\sqrt{\frac{2}{3}}P\hbar k_z & E_{LH}+\frac{\hbar^2k_z^2}{2m_0} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & E_{HH}+\frac{\hbar^2k_z^2}{2m_0} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{\sqrt{3}}P\hbar k_z & 0 & 0 & E_{SO}+\frac{\hbar^2k_z^2}{2m_0}
\end{pmatrix}. \tag{2.16}$$

The simplification of this approximation provides cannot be overstated. First of all, envelopes belonging to states of opposite spin have completely decoupled from each other. Secondly, the heavy-hole envelopes have decoupled from the conduction band, light-hole, and split-off envelopes, a result of the symmetry of the Bloch functions. The heavy-hole band is not generally of interest in QCLs, and there is no explicit spin-dependence in the matrix, so what was once an eight-band eigenvalue problem has reduced to a single three-band one. Using this more compact three-component notation, and replacing the diagonal energy terms with their material- and potential-dependent counterparts $E_C(z) \equiv E_C^{(A)}U_A + E_C^{(B)}U_B + V_{ext}$, $E_{LH}(z) \equiv E_{LH}^{(A)}U_A + E_{LH}^{(B)}U_B + V_{ext}$, and $E_{SO}(z) \equiv E_{SO}^{(A)}U_A + E_{SO}^{(B)}U_B + V_{ext}$, the envelope Schrödinger equation reduces to

$$\begin{pmatrix} -\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial z^2} + E_C(z) & i\sqrt{\frac{2}{3}}P\hbar\frac{\partial}{\partial z} & -i\sqrt{\frac{1}{3}}P\hbar\frac{\partial}{\partial z} \\ i\sqrt{\frac{2}{3}}P\hbar\frac{\partial}{\partial z} & -\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial z^2} + E_{LH}(z) & 0 \\ -i\sqrt{\frac{1}{3}}P\hbar\frac{\partial}{\partial z} & 0 & -\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial z^2} + E_{SO}(z) \end{pmatrix} \begin{pmatrix} F_C \\ F_{LH} \\ F_{SO} \end{pmatrix} = E \begin{pmatrix} F_C \\ F_{LH} \\ F_{SO} \end{pmatrix}. \tag{2.17}$$

The so-called free electron terms $\frac{\hbar^2 k_z^2}{2m_0}$ can be shown to be of order $(E_C - E_{LH})/E_p$, where $E_p \equiv 2m_0P^2$ is called the Kane energy. [11] Since $E_p \approx 20\,eV$ and $E_C - E_{LH} \approx 0.3\,eV$ for most materials, those terms can generally be neglected. For states with energies above the conduction band, one can then write down the second and third rows of (2.17) directly as

$$F_{LH} = +iP\hbar\sqrt{\frac{2}{3}}\frac{1}{E - E_{LH}(z)}\frac{\partial}{\partial z}F_C \tag{2.18}$$

$$F_{SO} = -iP\hbar\sqrt{\frac{1}{3}}\frac{1}{E - E_{SO}(z)}\frac{\partial}{\partial z}F_C. \tag{2.19}$$

Plugging these back into the first row of (2.17), one finds that F_C must obey the relation

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\frac{1}{m_0}\left[\frac{2}{3}\frac{E_p}{E - E_{LH}(z)} + \frac{1}{3}\frac{E_p}{E - E_{SO}(z)}\right]\frac{\partial}{\partial z}F_C + E_C(z)F_C = EF_C. \tag{2.20}$$

The quantity in between the derivatives has units of inverse mass, and it is useful to define the effective mass $m^*(E,z) \equiv m_0\left[\frac{2}{3}\frac{E_p}{E-E_{LH}(z)} + \frac{1}{3}\frac{E_p}{E-E_{SO}(z)}\right]^{-1}$, in which case (2.20) becomes

$$\boxed{\left(-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\frac{1}{m^*(E,z)}\frac{\partial}{\partial z} + E_C(z)\right)F_C = EF_C.} \tag{2.21}$$

Thus, Schrödinger's equation has reduced to a relatively simple one-dimensional eigenvalue equation, albeit one where the effective mass is energy- and position-dependent. Its simple form allows it to be solved numerically, but because the problem is not a Hermitian one, great care must be taken not to assign properties such as orthonormality to its solutions.

2.4 Optical transition rates

When measuring the properties of a QCL structure, the main quantity of interest will be the absorptivity of the electromagnetic modes propagating through the QCL waveguide, and so it is essential that theses processes be well-understood. Even though the waveguide modes are in general not plane waves, the coupling between them and the electrons in the system is quite weak, and it is sufficient to model them as such. In a region with constant magnetic permeability and effective modal index n, a plane wave of intensity I, angular frequency ω, and unit polarization vector $\hat{e}_{\mathbf{k},\sigma}$ has a vector potential of the form

$$\mathbf{A} = \sqrt{\frac{\mu_0 c I}{2\omega^2 n}} \hat{e}_{\mathbf{k},\sigma} \left(e^{i(k\cdot\mathbf{r}-\omega t+\phi)} + e^{-i(k\cdot\mathbf{r}r-\omega t+\phi)} \right). \tag{2.22}$$

In the Coulomb gauge, defined by $\nabla \cdot \mathbf{A} = 0$, the scalar potential ϕ vanishes, and the operators $\mathbf{A}$ and $\mathbf{p}$ commute. Ignoring spin-orbit coupling, the electromagnetic Hamiltonian is found by replacing $\mathbf{p}$ in Schrödinger's equation with $\mathbf{p}-e\mathbf{A}$ and adding a scalar potential term $-e\phi(\mathbf{r})$. [12] The full Hamiltonian then becomes

$$\begin{aligned} \mathcal{H} &= \frac{1}{2m_0}(\mathbf{p} - e\mathbf{A})^2 + V(\mathbf{r}) - e\phi \\ &= \frac{\mathbf{p}^2}{2m_0} - \frac{e}{2m_0}(\mathbf{A}\cdot\mathbf{p} + \mathbf{p}\cdot\mathbf{A}) + \frac{e^2\mathbf{A}^2}{2m_0} + V(\mathbf{r}) \\ &= \mathcal{H}_0 - \frac{e}{m_0}\mathbf{A}\cdot\mathbf{p} + \frac{e^2\mathbf{A}^2}{2m_0}. \end{aligned} \tag{2.23}$$

Both electromagnetic terms in the Hamiltonian are small, but the second one corresponds to two-photon absorption and can usually be neglected. To find the transition rate of electrons between an initial state $|\mathrm{i}\rangle$ and a final state $|\mathrm{f}\rangle$, one should use the results of time-dependent perturbation theory, which simplifies to Fermi's Golden Rule in the case of harmonic perturbation: [12]

$$W_{\mathrm{i}\to\mathrm{f}} = \frac{2\pi}{\hbar}\left[\left|\left\langle \mathrm{f}\left|\mathcal{H}'_{-}\right|\mathrm{i}\right\rangle\right|^2 \delta(E_{\mathrm{f}} - E_{\mathrm{i}} - \hbar\omega) + \left|\left\langle \mathrm{f}\left|\mathcal{H}'_{+}\right|\mathrm{i}\right\rangle\right|^2 \delta(E_{\mathrm{f}} - E_{\mathrm{i}} + \hbar\omega)\right]. \tag{2.24}$$

Inserting (2.22) and (2.23) into (2.24), and exploiting the fact that the typical wavefunction is much smaller than the length scale of the field, the $e^{i(k\cdot\mathbf{r}+\phi)}$ and $e^{-i(k\cdot\mathbf{r}+\phi)}$ terms become simple phase factors and vanish, and one finds that

$$W_{\mathrm{i}\to\mathrm{f}} = \frac{\pi e^2 \mu_0 c I}{\hbar m_0^2 \omega^2 n}\left|\langle \mathrm{f}\left|\hat{\mathrm{e}}_{\mathbf{k},\sigma}\cdot\mathbf{p}\right|\mathrm{i}\rangle\right|^2\left[\delta(E_{\mathrm{f}} - E_{\mathrm{i}} - \hbar\omega) + \delta(E_{\mathrm{f}} - E_{\mathrm{i}} + \hbar\omega)\right]. \tag{2.25}$$

Up until now, complete generality has been assumed for the forms of the states $|\mathrm{i}\rangle$ and $|\mathrm{f}\rangle$. If one assumes that transitions arise from the subbands of a QCL, then those states can be taken to be of the form

$$\begin{aligned} |\mathrm{i}\rangle &= \frac{1}{\sqrt{A}}\, e^{i\mathbf{k}_\perp^{(i)}\cdot\mathbf{r}} \sum_n F_n^{(i)}(z)\vec{u}_{n0}(\mathbf{r}) \\ |\mathrm{f}\rangle &= \frac{1}{\sqrt{A}}\, e^{i\mathbf{k}_\perp^{(f)}\cdot\mathbf{r}} \sum_m F_m^{(f)}(z)\vec{u}_{m0}(\mathbf{r}). \end{aligned} \tag{2.26}$$

where the $\mathbf{k}_\perp$'s represent the transverse momentum and the summations are taken for now over all eight bands. Using a process similar to that of (2.14), one finds that the momentum matrix element is given by

$$\begin{aligned} \langle \mathrm{f}\,|\,\mathbf{p}|\mathrm{i}\rangle &= \sum_{n,m}\int d^3\mathbf{r}\frac{1}{A}\, e^{i(\mathbf{k}_\perp^{(i)}-\mathbf{k}_\perp^{(f)})\cdot\mathbf{r}} F_m^{(f)^*}(z)\left[\hbar\mathbf{k}_\perp^{(i)} + \hat{z}p_z + \langle u_{m0}\,|\,\mathbf{p}|u_{n0}\rangle\right] F_n^{(i)}(z) \\ &= \delta_{k_\perp^{(i)},k_\perp^{(f)}}\int\sum_{n,m} F_m^{(f)^*}(z)\left[\hbar\mathbf{k}_\perp^{(i)} + \hat{z}p_z + \langle u_{m0}\,|\,\mathbf{p}|u_{n0}\rangle\right] F_n^{(i)}(z)\, dz. \end{aligned} \tag{2.27}$$

The ramifications of this statement are twofold. For one thing, states having different transverse momenta cannot couple to one another, due to conservation of momentum. Additionally, for k_x and k_y near zero, any terms pointing in the $\hat{x}$ and $\hat{y}$ direction will vanish. Therefore, TM-polarized light—light polarized in the growth direction—

will see the intersubband transitions, while TE-polarized light—light polarized in the transverse direction—will see bulk material. This is the polarization selection rule, and is at the heart of any absorptivity measurement.

From now on, consider only TM-polarized light, for which $\hat{e}_{\mathbf{k},\sigma} = \hat{z}$. The transition rate for a single pair of states can be written as

$$W_{\mathrm{i}\to\mathrm{f}} = \frac{\pi e^2 \mu_0 c I}{\hbar m_0^2 \omega^2 n}\, \delta_{k_\perp^{(i)},k_\perp^{(f)}} \left|\langle \psi_\mathrm{f}\left|p_z\right|\psi_\mathrm{i}\rangle\right|^2 \left[\delta(E_\mathrm{f} - E_\mathrm{i} - \hbar\omega) + \delta(E_\mathrm{f} - E_\mathrm{i} + \hbar\omega)\right], \quad (2.28)$$

where $\langle \psi_\mathrm{f}\left|p_z\right|\psi_\mathrm{i}\rangle \equiv \int \sum_{n,m} F_m^{(f)^*}(z)\,(p_z + \langle u_{m0}\,|\,p_z|u_{n0}\rangle)\,F_n^{(i)}(z)\,dz$ is the reduced momentum matrix element—to be evaluated later. Assuming that the transverse effective masses for different subbands are equal, and that $E_\mathrm{i} < E_\mathrm{f}$, the joint density of states between them will be delta function-like, and the total per area transition rate from subband i to subband f will be

$$\frac{r_{\mathrm{i}\to\mathrm{f}}}{A} = \frac{\pi e^2 \mu_0 c I N_\mathrm{i}}{\hbar m_0^2 \omega^2 n} \left|\langle \psi_\mathrm{f}\left|p_z\right|\psi_\mathrm{i}\rangle\right|^2 \delta(E_\mathrm{f} - E_\mathrm{i} - \hbar\omega) \quad (2.29)$$

where N_i is the sheet density of electrons in subband i. This process of scattering of electrons from i to f essentially an upward absorption process, but there will also be a simultaneous process arising from stimulated emission of electrons f to i occurring at the same rate. Thus, the total rate at which the field loses energy will have contributions from both, weighted by the photon energy $\hbar\omega$:

$$\frac{1}{A}\frac{dE}{dt} = -\frac{\pi e^2 \mu_0 c I (N_\mathrm{i} - N_\mathrm{f})}{m_0^2 \omega n} \left|\langle \psi_\mathrm{f}\left|p_z\right|\psi_\mathrm{i}\rangle\right|^2 \delta(E_\mathrm{f} - E_\mathrm{i} - \hbar\omega) \quad (2.30)$$

Assuming that the field loses energy uniformly in the $\hat{x}$ and $\hat{y}$ plane, but only over the span of a single QCL module in the $\hat{z}$-direction, the change in energy density over time inside that module is given by

$$\frac{d\rho_E}{dt} = \frac{1}{AL_{mod}}\frac{dE}{dt} = -\frac{\pi e^2 \mu_0 c I}{m_0^2 \omega n}\frac{N_\mathrm{i} - N_\mathrm{f}}{L_{mod}} \left|\langle \psi_\mathrm{f}\left|p_z\right|\psi_\mathrm{i}\rangle\right|^2 \delta(E_\mathrm{f} - E_\mathrm{i} - \hbar\omega) \quad (2.31)$$

A plane wave traveling in the $\hat{y}$-direction at a speed c/n has an intensity $I = \frac{c}{n}\rho_E$, and travels an infinitesimal distance $dy = \frac{c}{n}dt$ in the time dt. Putting these together, one can formally write $\frac{dI}{dy} = \frac{n}{c}\frac{dI}{dt} = \frac{d\rho}{dt}$, or

$$\frac{dI}{dy} = -\left(\frac{\pi e^2 \mu_0 c}{m_0^2 \omega n}\frac{N_\mathrm{i} - N_\mathrm{f}}{L_{mod}} \left|\langle \psi_\mathrm{f} \left| p_z \right| \psi_\mathrm{i} \rangle\right|^2 \delta(E_\mathrm{f} - E_\mathrm{i} - \hbar\omega)\right) I \equiv -\alpha I. \tag{2.32}$$

Because the energy density was assumed to be lost only in the span of a single module, the above absorptivity only applies within that span, and the modal absorption coefficient must be found by summing this over *all* modules. Fortunately, all modules are identical and the absorption region for each module was defined to be non-overlapping, so if the modal confinement factor Γ is defined to be the fraction of the mode's integrated power density that overlaps the active region, the modal absorption will be simply given by

$$\alpha(\hbar\omega) = \Gamma \frac{\pi e^2 \mu_0 c \Delta N}{m_0^2 \omega n} \left|\langle \psi_\mathrm{f} \left| p_z \right| \psi_\mathrm{i} \rangle\right|^2 \delta(E_\mathrm{f} - E_\mathrm{i} - \hbar\omega), \tag{2.33}$$

where the sheet density terms have been replaced with an average volume density $\Delta N \equiv \frac{N_\mathrm{i} - N_\mathrm{f}}{L_{mod}}$. To express this instead in terms of frequency, one defines $\nu_0 \equiv (E_\mathrm{f} - E_\mathrm{i})/(2\pi\hbar)$ and $\nu \equiv \omega/(2\pi)$, obtaining

$$\alpha(\nu) = \Gamma \frac{\Delta N e^2 \mu_0 c}{4\pi\hbar m_0^2 n \nu_0} \left|\langle \psi_\mathrm{f} \left| p_z \right| \psi_\mathrm{i} \rangle\right|^2 \gamma(\nu - \nu_0), \tag{2.34}$$

where the delta function has also been replaced with a linewidth function $\gamma(\nu)$ in order to reflect the fact that other mechanisms such as nonparabolicity and interface scattering will broaden the linewidth. (By convention, $\gamma(\nu)$ is normalized so that $\int \gamma(\nu)\, d\nu = 1$.) Lastly, to bridge the notion of quantum mechanical transitions and classical Hertzian dipoles, it is sometimes useful to define the unscaled, dimensionless oscillator strength. Letting $f_{\mathrm{i}\to\mathrm{f}} \equiv \frac{2}{m_0}\frac{\left|\langle \psi_\mathrm{f} \left| p_z \right| \psi_\mathrm{i} \rangle\right|^2}{2\pi\hbar\nu_0}$, the absorption reduces to

$$\alpha(\nu) = \Gamma \frac{\Delta N e^2 \mu_0 c f_{\mathrm{i}\to\mathrm{f}}}{4 m_0 n} \gamma(\nu - \nu_0) \tag{2.35}$$

Note that regardless of the actual shape of the linewidth function, integrating (2.35) over all frequencies gives a result proportional to the oscillator strength of the two states and to the population difference. If either can be known to a high degree of certitude, and the total absorption spectrum can be measured, then the other can be found.

2.5 Reduced momentum matrix element

To complete the analysis of the optical absorptivity, the reduced matrix element $\langle \psi_{\mathrm{f}} \left| p_z \right| \psi_{\mathrm{i}} \rangle$ still needs to be evaluated. For one-dimensional states, the double-summation only needs to be evaluated over the conduction, light-hole, and split-off bands:

$$\langle \psi_{\mathrm{f}} \left| p_z \right| \psi_{\mathrm{i}} \rangle = \int \sum_{n,m} F_m^{(f)^*}(z) \left(p_z + \langle u_{m0} \left| p_z \right| u_{n0} \rangle \right) F_n^{(i)}(z) \, dz. \tag{2.36}$$

Fortunately, the Bloch momentum matrix elements can be found from their analytical forms in Table 2.1, and can be succinctly described in their matrix form as

$$\overline{\overline{p_z}} = \begin{pmatrix} 0 & -\sqrt{\frac{2}{3}} P m_0 & \sqrt{\frac{1}{3}} P m_0 \\ -\sqrt{\frac{2}{3}} P m_0 & 0 & 0 \\ \sqrt{\frac{1}{3}} P m_0 & 0 & 0 \end{pmatrix} \tag{2.37}$$

Because the free-electron terms were dropped from (2.17) in order to derive the effective mass Hamiltonian, the standalone p_z term also needs to be dropped in order to maintain consistency. [11] In that case, the matrix element becomes

$$\begin{aligned} \langle \psi_{\mathrm{f}} \left| p_z \right| \psi_{\mathrm{i}} \rangle &= \int F_C^{(\mathrm{f})^*} \left(-\sqrt{\frac{2}{3}} P m_0 F_{LH}^{(\mathrm{i})} + \sqrt{\frac{1}{3}} P m_0 F_{SO}^{(\mathrm{i})} \right) \\ &\qquad + \left(-\sqrt{\frac{2}{3}} P m_0 F_{LH}^{(\mathrm{f})} + \sqrt{\frac{1}{3}} P m_0 F_{SO}^{(\mathrm{f})} \right) F_C^{(\mathrm{i})^*} \, dz \\ \langle \psi_{\mathrm{f}} \left| p_z \right| \psi_{\mathrm{i}} \rangle &= \frac{1}{2} \int F_C^{(\mathrm{f})^*} \left(\frac{m_0}{m^*(E_{\mathrm{f}}, z)} p_z + p_z \frac{m_0}{m^*(E_{\mathrm{f}}, z)} \right) F_C^{(\mathrm{i})} \, dz. \end{aligned} \tag{2.38}$$

Once again, the microscopic dependence of the crystal has completely vanished, allowing the element to be calculated from the conduction band envelope function (which in turn obeys the effective mass Schrödinger equation). It should be mentioned that in the limit of no conduction band nonparabolicity, the oscillator strength can alternatively be written in terms of the one-band dipole moment: [13]

$$f_{\mathrm{i}\to\mathrm{f}} = \frac{2m_0(E_\mathrm{f} - E_\mathrm{i})}{\hbar^2} \left| \left\langle F_C^{(\mathrm{f})} \left| z \right| F_C^{(\mathrm{i})} \right\rangle \right|^2 . \tag{2.39}$$

Though this definition is simpler, it is also less accurate, and requires that the envelope functions be orthonormal to each other. The ramifications of using it will be discussed in Chapter 3.

Chapter 3

Numerical simulation

A large variety of factors affect the performance and absorptivity of QCLs, and so numerical simulation is essential for an accurate description of their behavior. These effects can primarily be classified into two categories: those related to the subband structure and those related to the electromagnetic properties of the waveguide. Both categories will be discussed here.

3.1 Quantum mechanical simulations

3.1.1 Material parameters and strain effects

Obviously, the important properties of a QCL will depend greatly on the parameters of its component materials. Strictly speaking, alloys of different III-V materials are not periodic crystals, meaning that they cannot be treated using the formalism developed in Chapter 2. However, it is generally considered sufficient to treat them as such through the use of the virtual crystal approximation, in which the material parameters take the form of simple quadratic functions of the alloy concentration. For consistency's sake, all material parameters used in this work were obtained from the review compiled by Vurgaftman et al. [14] It must be noted that some parameters have experimental errors as high as 25%. This will have important consequences when agreement is sought between measured and simulated data, because it means that some degree of leeway should be allowed in order to match the data properly. This particularly applies to the well material's effective mass, which holds great influence over the energy of all transitions. (Since wavefunctions are reduced within the barriers, transition energies are less sensitive to the barrier mass and conduction band offset.)

As should be clear from Chapter 2, the main parameters of interest are E_p, E_C, E_g, and Δ (or equivalently, P, E_C, E_{LH}, and E_{SO}). Note that the bulk effective mass was not included, for the simple reason that it is not independent of those four parameters: the tabulated effective mass is related to the Kane energy by evaluating the effective mass defined previously at $E = E_C$, or

$$m^* = m_0 \left[\frac{2}{3} \frac{E_p}{E_g} + \frac{1}{3} \frac{E_p}{E_g + \Delta} \right]^{-1} \tag{3.1}$$

Unfortunately, finding the other band structure parameters is not as straightforward a process. Because the materials under consideration are typically highly strained, the Bloch functions are changed from what they are in bulk material, as are the band offsets. Making matters worse, the atomic potential at the interface of two materials could be different from the potential in the center, meaning that a full description of strain should technically undo most of the 8×8 $\mathbf{k} \cdot \mathbf{p}$ envelope function analysis.

Nonetheless, Van de Walle managed to create a model-solid theory that effectively describes the behavior of strained materials. Like Kane's formulation, model-solid theory uses the results of a complicated problem—in this case, local-density-functional and pseudopotential calculations—and lumps most of that behavior into a few tabulated parameters. [15] First, he defines the in-plane strain as

$$\varepsilon = \varepsilon_{xx} = \varepsilon_{yy} \equiv \frac{a_0 - a(x)}{a_0} \tag{3.2}$$

where a_0 and $a(x)$ are the respective lattice constants of the substrate and the material. Next, the conduction, average valence band, and auxiliary shifts are defined as

$$\delta E_C \equiv 2a_c(1-\frac{C_{12}}{C_{11}})\varepsilon \tag{3.3}$$

$$\Delta E_{\mathrm{v,avg}} \equiv 2a_v(1-\frac{C_{12}}{C_{11}})\varepsilon \tag{3.4}$$

$$\delta E_{001} \equiv -2b(1+2\frac{C_{12}}{C_{11}})\varepsilon \tag{3.5}$$

where a_c, a_v, b, C_{12}, and C_{11} are the tabulated parameters that lump together the effects of strain on the material. Continuing, the heavy-hole, light-hole, and split-off splittings are defined as

$$\Delta E_{HH} \equiv \frac{1}{3}\Delta - \frac{1}{2}\delta E_{001} \tag{3.6}$$

$$\Delta E_{LH} \equiv -\frac{1}{6}\Delta + \frac{1}{4}\delta E_{001} + \frac{1}{2}\sqrt{\Delta^2 + \Delta\,\delta E_{001} + \frac{9}{4}(\delta E_{001})^2} \tag{3.7}$$

$$\Delta E_{SO} \equiv -\frac{1}{6}\Delta + \frac{1}{4}\delta E_{001} - \frac{1}{2}\sqrt{\Delta^2 + \Delta\,\delta E_{001} + \frac{9}{4}(\delta E_{001})^2} \tag{3.8}$$

At this point, the final positions of all four bands are found on an *absolute* scale, through the use of the tabulated $E_{v,\mathrm{avg}}$ parameter (which represents the average energy position of the three valence bands). This is extremely important, because it allows the same energy scale to be used over *all* materials. It is in turn is shifted by the strain-dependent quantity $\Delta E_{\mathrm{v,avg}}$, and the final band positions are

$$E_C = E_{\mathrm{v,avg}} + \Delta E_{\mathrm{v,avg}} + \frac{\Delta}{3} + E_g + \delta E_C \tag{3.9}$$

$$E_{LH} = E_{\mathrm{v,avg}} + \Delta E_{\mathrm{v,avg}} + \Delta E_{LH} \tag{3.10}$$

$$E_{HH} = E_{\mathrm{v,avg}} + \Delta E_{\mathrm{v,avg}} + \Delta E_{HH} \tag{3.11}$$

$$E_{SO} = E_{\mathrm{v,avg}} + \Delta E_{\mathrm{v,avg}} + \Delta E_{SO} \tag{3.12}$$

3.1.2 Shooting method solver

As seen in Chapter 2, the complicated problem of solving Schrödinger's equation simplifies for heterostructures into a relatively simple eigenvalue equation:

$$\left(-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\frac{1}{m^*(E,z)}\frac{\partial}{\partial z}+E_C(z)\right)F_C=EF_C \tag{3.13}$$

$$m^*(E,z)\equiv m_0\left[\frac{2}{3}\frac{E_p}{E-E_{LH}(z)}+\frac{1}{3}\frac{E_p}{E-E_{SO}(z)}\right]^{-1} \tag{3.14}$$

Assuming that the values of E_C, E_{LH}, and E_{SO} are known across the simulation space, the goal will be to find the wavefunctions of the system and their corresponding energies. The simulation space is usually taken to be a few QCL modules, and vanishing wavefunction conditions are applied at the boundaries. (This choice is largely irrelevant, as only wavefunctions near the boundary are affected by this choice.)

At a given energy E, Schrödinger's equation is just a second-order ordinary differential equation in position, meaning that only two initial conditions (e.g., an initial value and slope) are required to describe its value at *all* locations. Since the overall normalization of the wavefunction is arbitrary, one can in principle apply the $F_C = 0$ condition at the left boundary, set the slope $\frac{\partial F_C}{\partial z}$ at the left boundary to some arbitrary fixed value, and find the wavefunction value at the right boundary. Even though this can be done for all real values of E, only eigenenergies will satisfy the additional boundary condition that this final value vanishes. In other words, if one were to define f(E) as the function that has energy as the input and the final value of the wavefunction as the output, finding the energy eigenvalues of the system corresponds to finding the zeros of f(E). Once these eigenvalues are found, finding the wavefunctions is just a trivial matter of integrating (3.13) across the module.

Exactly how f(E) is calculated is a matter of preference. One straightforward method is to use a discrete derivative to perform the integration, as described by Lundstrom. [16] This method is particularly fast for systems with applied electric fields, but is not very efficient under flat-band conditions. The method used for the present work was that of propagation matrices, in which the potential and mass terms are held constant over small discrete regions and the analytical form of the wavefunction is found in each region by matching boundary conditions. [12] To first order these methods are equivalent, but the latter has fewer discretization artifacts and is much faster under flat-band conditions, which comprised the majority of simulations.

It should be noted that even though the shooting method is very fast compared to other methods that can deal with band nonparabolicity, it is not very robust. For example, if a very small anticrossing exists in a structure, several roots of $f(\mathrm{E})$ will be very close to each other, and root-finding algorithms may not find one or more. For this reason, it is actually more useful to count the number of zero crossings *of the wavefunction*, and use this as a way to place bounds on the location of the roots. As energy is increased continuously from below an eigenvalue to above an eigenvalue, the number of zero crossings changes abruptly as an extra one is "pulled in" from the region outside the simulation space. Essentially, this provides a count for how many roots exist below any given energy value, and turns the hard problem of searching for roots into a deterministic one.

Another important issue is that even once the root locations have been found quite accurately, they might still be sufficiently far from the true root that the wavefunction appears to blow up at the final boundary, by virtue of the finiteness of $f(\mathrm{E})$. In principle, this can be solved by finding the root more accurately, but in practice, one finds that the root can be known to within computation precision and the wavefunction will still blow up! When this happens, the solution is to choose a nearby energy value, construct its corresponding wavefunction, and add the two wavefunctions weighted by coefficients that force the boundary value to vanish. Essentially, this exploits the fact that the wavefunction at all points can be expressed as linear functions of energy, and the superposition forms exactly the right weighted average to force the wavefunction to zero without disturbing its value over the rest of the space.

3.1.3 Self-consistent Poisson solver

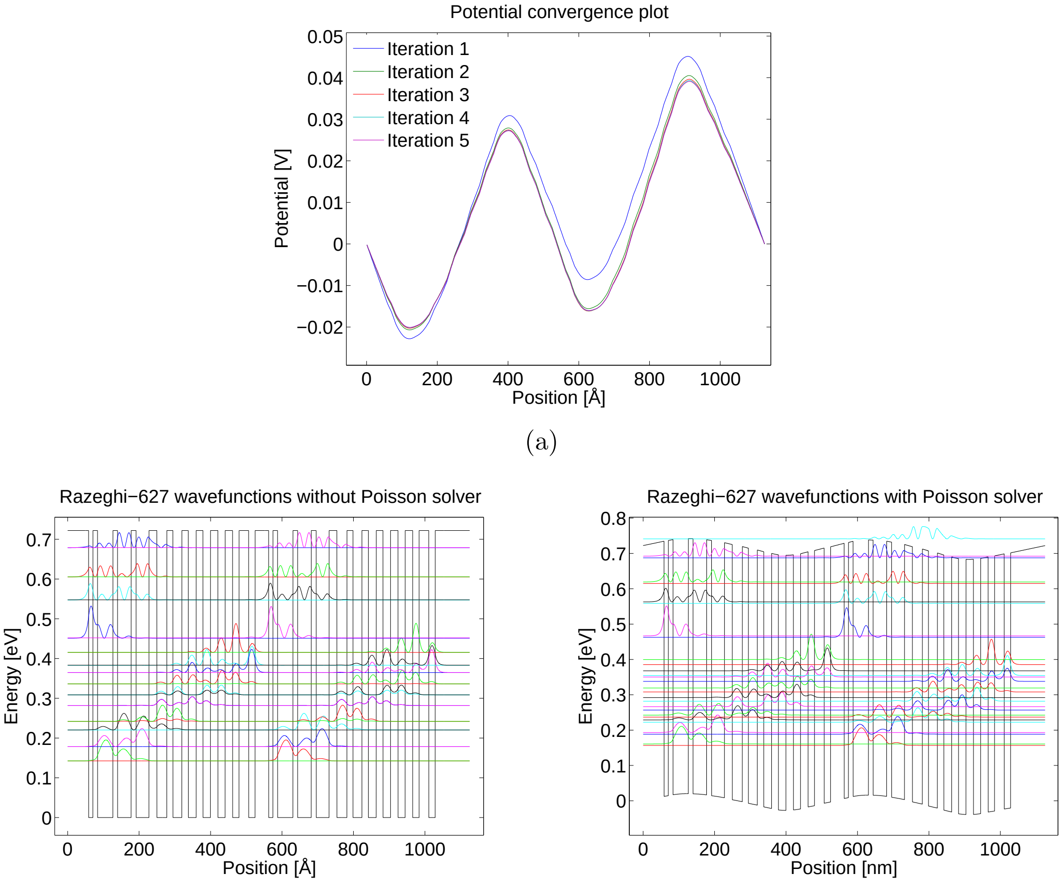
Because some of the QCL structures that have been fabricated have relatively high average doping densities, the relatively small effects of space charge on the wavefunctions cannot be neglected. In most structures, dopants are chosen to be as far from the lasing wavefunctions as possible, so as not to negatively impact the transition linewidth. Under design bias conditions, the effect of this relatively low, as the system's electrons tend to spread themselves out across the whole module. However, under equilibrium conditions, the effect is exaggerated, as the thermal distribution of electrons pools electrons in the lowest-lying states, far from the dopants. An internal field then develops, in which the potential obeys Poisson's equation in one dimension:

$$\frac{\partial}{\partial z}\left(\epsilon(z)\frac{\partial\phi}{\partial z}\right) = -\rho \tag{3.15}$$

where ϕ is the electric potential, ϵ is the material-dependent permittivity, and ρ is the charge density arising from both the positively-charged donors and the negatively-charged electrons. Clearly, once ϕ and ρ are known, a solution can quickly be found using propagation matrices. The only subtleties come from the boundary conditions of the system and from how the electron density is chosen. Under zero bias, the most sensible choice of boundary conditions is to force the potential and field to vary periodically across a module. To assign an electron charge density, it is assumed that wavefunctions' probability density corresponds directly to charge density—the so-called Hartree approximation—and that the electrons are distributed within subbands according to Boltzmann statistics.

Clearly, it is fairly easy to find the potential of a system given its band structure. Nonetheless, if an internal field has built up, electrons will move in such a way that they partially cancel it. In other words, the potential perturbs $E_C(z)$, forcing the wavefunctions and energies to be recalculated. Thus, finding the true band structure requires that Schrödinger's equation and Poisson's equation be solved *self-consistently.* To do this, an iterative process is used, in which Schrödinger's equation is first solved assuming no internal potential, followed by Poisson's equation, followed by the modified Schrödinger's equation, followed by Poisson's equation again, etc. As long as the doping is sufficiently low, this process will converge rapidly to the true solution. However, if the doping is sufficiently high, or if a large number of modules is simulated, it is possible for the simulation to diverge. When this occurs, convergence can be obtained by applying a damping parameter that lessens the speed at which the charge density is allowed to change.

Figure 3-1 shows the results of a two-module simulation of the Razeghi-427 series of devices, the highest-doped of the structures tested. Figure 3-1(a) shows the fast convergence of the iterative process; even with considerable damping applied, the electric potential becomes close to its final value after only two iterations. Figures 3-1(b) and 3-1(c) show the effects of using the Poisson solver. Note that while application of the Poisson solver does not appreciably change the shape of the wavefunctions, it has shifted their energies. However, at the same time, the energies corresponding to lasing transitions—those localized to the system's wells—have all shifted in the same direction, and so relatively little shift is observed in the overall transition energies.



(a)

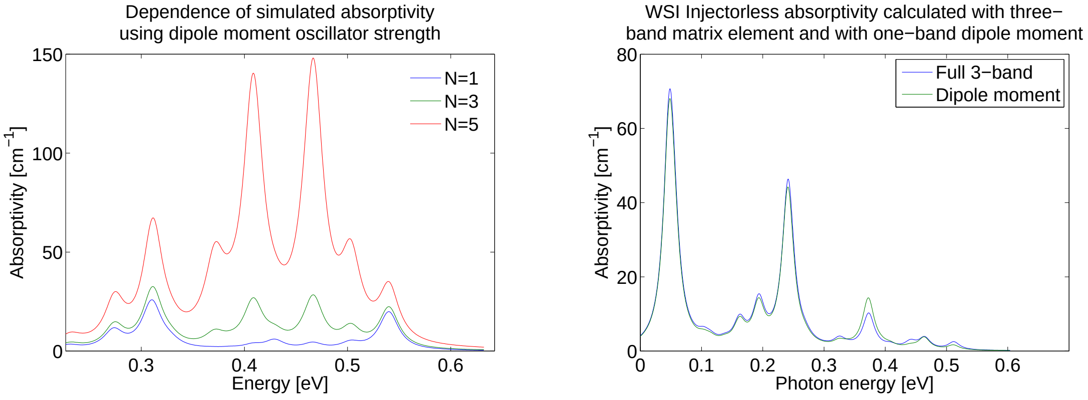
(b)

(c)

Figure 3-1: Schrödinger-Poisson simulations for Razeghi-427 structure (a) Convergence of electric potential (b) Wavefunctions without Poisson solver used (c) Wavefunctions with Poisson solver used

3.1.4 Absorption Calculation

Once the structure's wavefunctions and energies have been found by solving (3.13) and (3.15) self-consistently, finding the total absorptivity of the structure requires that the oscillator strengths between all pairs of states be calculated, using (2.38). While it is tempting to instead use the dipole moment formalism of (2.39), as it lacks a derivative and is therefore less sensitive to wavefunction discretization artifacts, its use cannot be justified for mid-infrared transitions. For one thing, if it is to be well-defined, the F_C envelopes must be orthonormal to each other, even though the non-Hermiticity of the envelope Schödinger equation implies that orthonormality is not guaranteed. Figure 3-2(a) demonstrates this issue by showing how the final computed absorptivity varies when the same structure is simulated with different numbers of modules. Analytically speaking, it should quickly converge to a fixed value, but since different states have slight overlaps with one another, it instead continues to increase. In addition, the correction provided by the correct expression is not always insignificant, as Figure 3-2(b) shows. Though the correction is small for transitions of low energy, about 5%, the correction for the highest energy transition is more than 40%.



(a) (b)

Figure 3-2: (a) Dependence of simulated absorptivity on number of modules with dipole moment formalism (b) Comparison of different methods for computing oscillator strength

Once the oscillator strengths have been properly calculated, finding the absorptivity is a simple matter of summing together the state-to-state absorptivities at all frequencies:

$$\alpha_{\mathrm{tot}}(\nu) = \sum_{\mathrm{i}} \sum_{\mathrm{j}>\mathrm{i}} \Gamma(\nu) \frac{\Delta N_{\mathrm{ij}}\, e^2 \mu_0 c f_{\mathrm{i}\to\mathrm{j}}}{4 m_0 n}\, \gamma(\nu - \nu_{0,\mathrm{ij}}). \tag{3.16}$$

In general, the mode confinement factor Γ will be frequency-dependent, and electromagnetic simulations are required to calculate this dependence. Furthermore, the population of each state must still be calculated. At equilibrium, because the doping is relatively low and the temperature is high, the average population of the ith state will be approximately given by Boltzmann statistics, or

$$N_{\mathrm{i}} = N_0 \exp(-\frac{E_{\mathrm{i}}}{kT}) \tag{3.17}$$

$$N_0 \equiv N_d \left(\sum_{\mathrm{i}} \exp(-\frac{E_{\mathrm{i}}}{kT}) \right)^{-1}, \tag{3.18}$$

where the constant N_0 was defined so that the average populations of all states add up to the total doping. This result only holds true at equilibrium: for devices under bias, more sophisticated techniques such as density-matrix Monte Carlo must be used. [17] [18]

The only remaining factor in the absorptivity that is related to the band structure is the shape of the linewidth function $\gamma(\nu)$. Because the linewidth depends on a multitude of factors, including growth quality, interface broadening, alloy broadening, band nonparabolicity, carrier lifetimes, etc., it is difficult to know *a priori* what the linewidth truly is. In fact, depending on homogeneity of the processes that contribute to the broadening, the shape may be Gaussian, Lorentzian, or something else entirely. If the shape is one of the former two, the linewidth function takes the form

$$\gamma_{\mathrm{G}}(\nu) = \frac{2\sqrt{\ln(2)}}{\Delta\nu\sqrt{\pi}} \exp\left(-\frac{4\ln(2)\nu^2}{\Delta\nu^2}\right) \tag{3.19}$$

$$\gamma_{\mathrm{L}}(\nu) = \frac{\frac{\Delta\nu}{2\pi}}{\nu^2 + \left(\frac{\Delta\nu}{2}\right)^2}, \tag{3.20}$$

where $\Delta\nu$ is the full-width-half-maximum (FWHM) of each function. Despite the impossibility of determining a linewidth for each possible optical transition, some general rules can be followed to produce plausible-looking absorption spectra. Firstly, the linewidths of high-energy transitions should be broader than the linewidths of low-energy transitions. For example, if the structure had some fixed "quality factor" associated with it, one would expect the linewidth to be roughly proportional to photon energy. Secondly, wavefunctions that overlap a large number of material boundaries should experience more broadening due to interface roughness. According to Wittman et al. [19], a structure with a fixed growth quality will have an interface roughness linewidth which obeys

$$\gamma_R \propto \sum_{z_k} \left(\left| F_C^{\mathrm{f}}(z_k) \right|^2 - \left| F_C^{\mathrm{i}}(z_k) \right|^2 \right)^2 \tag{3.21}$$

where the summation ranges over all interfaces in the system. It should be noted that these rules are not independent of one another, since states of high energy tend to be extended over many interfaces and vice versa. In general, linewidths are experimentally-determined parameters that can only be found after the fact.

3.2 Electromagnetic simulation

A thorough description of QCL absorption requires knowledge of its electromagnetic behavior as well as its band structure. Because broadband absorptivity was measured, these properties need to be described over a large wavelength range, from about 2 μm to 20 μm. Fortunately, most of the properties can be effectively encapsulated through just a few equations, which can be easily solved with the COMSOL numerical simulation package.

3.2.1 Two-dimensional mode solver

All of the measured QCLs have a waveguide structures that are typical of an index-guided Fabry-Perot laser. Essentially, a slab of high refractive index—the active region—is surrounded by two slabs of low refractive index—the cladding—and the whole structure is etched to provide lateral confinement. The top of the device is first coated in a thin layer of titanium, to serve as an adhesion promoter, and then in a thick layer of gold, which forms the electrical contact and also dissipates heat. The ends of the device are cleaved in such a way that a fraction of the light emitted from the laser into the waveguide will bounce back and forth inside the cavity due to Fresnel reflection, and the gain that the active region provides will permit lasing.

In order to enhance the mid-infrared performance of these devices, some modifications to this basic design were made. Instead of using just the active region for confinement, thin layers of inactive InGaAs layers (approximately 300 nm thick each) were grown between the active region and cladding. Because InGaAs has a much higher index than either the cladding or the active region, it has the effect of "pulling" the modes from the cladding and into the active region. Not only does this help to increase the mode confinement factor Γ, but it also helps to reduce the losses that arise from the interaction of the mode with the lossy titanium. As far as the lateral properties of the waveguide, they were generally processed into two varieties: ridges and broad-area devices. The ridge devices were the narrower of the two, having widths that ranged from 4 μm to 20 μm, while the broad areas had a fixed width of about 45 μm. Aside from width, they were also processed differently: after an ICP-RIE dry etch, the sidewalls of the ridges were coated in silicon dioxide to serve as an insulating layer, and the whole ridge was then coated in gold. By contrast, the broad area devices were wet-etched, had no sidewall coating of any kind, and only the top was coated in gold. Due to their superior heat-dissipation and fewer propagating modes, the ridges generally performed better. In addition, because the ridges did not require that wire bonds be placed directly in contact with the active region, they were less susceptible to the damage that could occur if a bond was placed improperly (e.g., a pulverized active region, or a peeling off of the gold). Thus, measurement of the ridge geometry was the primary concern.

Since the ridge geometry varies in both the lateral and growth directions on length scales similar to the wavelength, it is important that the two-dimensional modes of the waveguide be found and characterized. To that end, the standard waveguide analysis is performed, and the translational symmetry of the waveguide is exploited to write the electric and magnetic fields of the a single mode with propagation constant β as

$$\begin{aligned} \mathbf{E} &= e^{i\beta z}\mathbf{E}_\beta(x,y) \\ \mathbf{H} &= e^{i\beta z}\mathbf{H}_\beta(x,y) \end{aligned} \tag{3.22}$$

Note that by convention, $\hat{z}$ is chosen to point in the propagation direction, $\hat{y}$ points in the growth direction, and $\hat{x}$ points in the lateral direction. (This is in contrast to the band structure simulation, in which $\hat{z}$ is chosen to be the growth direction.) It is easy to show that if the magnetic permeability is constant ($\mu = \mu_0$) and the electric permittivity is only a function of x and y ($\epsilon = \epsilon_0\epsilon_r(x,y)$), Maxwell's equations reduce from their six-component form to a two-component eigenvalue problem:

$$\left[\epsilon_r\left(\frac{\omega}{c}\right)^2 + \vec{\nabla}_t^2 - \left(\vec{\nabla}\times\hat{z}\frac{1}{\epsilon_r}\right)\nabla_z\times\right]\mathbf{H}_t = \beta^2\mathbf{H}_t, \tag{3.23}$$

where $\mathbf{H}_t \equiv \hat{x}H_x + \hat{y}H_y$ contains the transverse variation of the magnetic field, $\nabla_z\times$ is the $\hat{z}$-component of the curl, and $\vec{\nabla}_t^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the two-dimensional vector Laplacian. Once the solutions of (3.23) have been found, the full field solutions can be written as

$$\begin{aligned} \mathbf{H}_\beta &= \mathbf{H}_t + \hat{z}\frac{i}{\beta}\vec{\nabla}\cdot\mathbf{H}_t \\ \mathbf{E}_\beta &= \frac{i}{\omega\epsilon_0\epsilon_r}\left[i\beta\hat{z} + \vec{\nabla}\right]\times\mathbf{H}_\beta. \end{aligned} \tag{3.24}$$

Using the regular machinery of vector spaces, one can show [20] that modes with different eigenvalues can be made orthonormal, in the sense that

$$\hat{z}\cdot\frac{1}{4}\int\left(\mathbf{E}_\beta^*\times\mathbf{H}_{\beta'} + \mathbf{E}_{\beta'}\times\mathbf{H}_\beta^*\right)\,d^2\mathbf{r}_t = \delta_{\beta',\beta}. \tag{3.25}$$

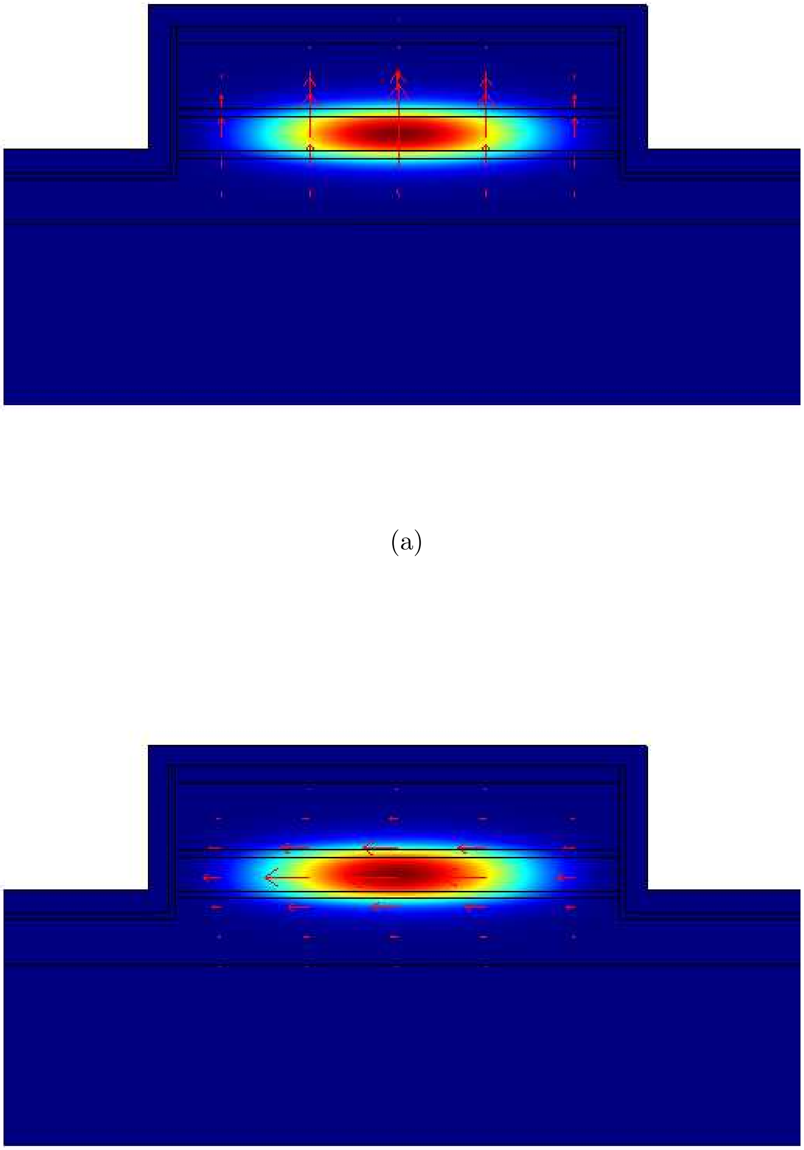
Notice that when $\beta = \beta'$, the quantity on the left hand side reduces to the total power flowing in the $\hat{z}$-direction. This will have useful consequences when calculating the frequency- and polarization-dependent coupling efficiencies.

Because (3.23) is a two-dimensional eigenvalue problem that is in general quite complicated, its solutions were obtained numerically using the COMSOL simulation package. First, the structure's geometry was analyzed, and the core values of the permittivity were determined in each region using a consistent set of material parameters. [21] As before, the material parameters for ternary compounds were determined through interpolation, but since no data was available for the quadratic bowing parameter, a simple linear interpolation was instead used. Furthermore, because the electromagnetic mode varies much more slowly than any of the wavefunctions in the active region, the active region can be accurately modeled as a region of constant permittivity, in which it is given simply by its average. Next, the complex permittivity in each region was found by an application of the Drude model, [13] which assumes that the permittivity takes the form

$$\epsilon(\omega) = \epsilon_{\text{core}} \left(1 - \frac{\omega_p^2 \tau^2}{1 + (\omega\tau)^2} + i \frac{\omega_p^2 \tau}{\omega(1 + (\omega\tau)^2)} \right), \tag{3.26}$$

where $\omega_p \equiv ne^2/m^* \epsilon_{\text{core}}$ is the plasma frequency of each material and τ is the Drude scattering time (determined for the semiconductors by their mobilities, determined for the metals by tabulation). The only exception is in the lossy SiO_2 layers of the ridges, whose complex permittivity is considered to have no frequency dependence. Lastly, COMSOL solves for the modes of the structure using a finite element method, in which the mesh size is roughly determined by the wavelength of light in the local material.

Figure 3-3 shows the results of a COMSOL simulation for a 20 μm ridge structure at a wavelength of 5 μm. Plotted are the two lowest-order modes: the color gradient represents power flow as a function of position, and the arrows represent the direction of the electric field. Notice that each solution is polarized purely in one dimension, reflecting the waveguide's pure TE and pure TM modes. As is the usual convention, TM refers to those modes polarized in the growth direction, and TE refers to modes polarized in the lateral direction. In addition, note that because the waveguide ridge was chosen to be relatively wide, a large number of lateral modes will appear in device operation. Had a smaller ridge been chosen, there may have been as few as



(a)

(b)

Figure 3-3: Fundamental mode profiles of 20 μm ridge structure at λ= 5 μm (a) TM mode (b) TE mode two modes—one TE and one TM—propagating in the structure.

Having determined the value of the propagation constant β for each mode, its behavior can be characterized in terms of two parameters that make the mode analogous to a plane wave: an effective waveguide index and a *power* waveguide loss parameter. Clearly, if one identifies them as $n_{\text{eff}} \equiv \frac{c}{\omega}\text{Re}(\beta)$ and $\alpha_w \equiv 2\text{Im}(\beta)$, the field's $\hat{z}$-dependence takes the form

$$e^{i\beta z} = e^{i\frac{\omega}{c}n_{\text{eff}}z}e^{-\frac{1}{2}\alpha_w z}, \tag{3.27}$$

which is superficially similar to the dependence of a plane wave traveling in the $+\hat{z}$ direction. In addition, the mode confinement factor Γ can be calculated by finding the fraction of power flow that occurs in the active region, i.e.,

$$\Gamma = \frac{\int_{\text{active}} \hat{z} \cdot \text{Re}\left(\mathbf{E}_\beta^* \times \mathbf{H}_\beta\right) \mathrm{d}^2\mathbf{r}_{\mathrm{t}}}{\int_{\text{space}} \hat{z} \cdot \text{Re}\left(\mathbf{E}_\beta^* \times \mathbf{H}_\beta\right) \mathrm{d}^2\mathbf{r}_{\mathrm{t}}}. \tag{3.28}$$

Because the modes of the system are well-confined near the design wavelength, with $\Gamma \approx 0.6$, the effective index is typically similar to the index of the active region, and $n_{\text{eff}} \approx 3.3$. The modes are also supposed to be low-loss in this regime, with simulated losses of about $\alpha_w \approx 1$ cm^{-1}. While n_{eff} is relatively frequency-independent, Γ and α_w can change drastically over a broad frequency range. Towards long wavelengths (above 10 μm), modes are weakly bound to the active region, increasing their losses. Towards short wavelengths (below 2 μm), modes are pulled into the high-index InGaAs waveguide layers, keeping their losses low but also lowering their confinement factor. These effects must be accounted for in any broadband absorption measurement.

Lastly, it should be mentioned that as long as the wavelength under consideration is below about 10 μm, there is a near-degeneracy that occurs between different polarizations in all of the structures simulated. Though the propagation constants are not exactly the same, their losses, confinement factors, and effective indices are quite similar. In addition, their mode profiles are practically identical, as Figure 3-3 shows. Ultimately, this happens because both polarizations are index-guided in the growth direction, and since all of the power is reflected from the cladding, the only effect changing the polarization could have is to change the decay length of the field in the cladding. This changes the confinement factor a little bit, but has a completely negligible effect on the loss and effective index.

3.2.2 Coupling efficiency calculation

The coupling of light to and from laser waveguides is another important effect that must not be misconstrued if a definitive measurement is to be made. Typically, one assumes that the coupling angle is small, in which case a scalar wave theory describes how externally-impinged light excites waveguide modes. In this case, the far field coupling efficiency essentially reduces to the Fourier transform of the power flow diagrams shown in Figure 3-3. [22] Unfortunately, this analysis does *not* apply to systems in which the coupling angle is significant, and since high numerical-aperture objectives were used to perform transmission measurements, a full vector wave simulation must be performed.

First, coupling into the waveguide is examined. The geometry of the problem will be such that the region $z < 0$ consists of free space, while the region $z > 0$ consists of an already-simulated waveguide. As in the scalar wave case, only far-field effects are considered, and incoming waves are assumed to be plane waves of the form

$$\begin{aligned} \mathbf{E}_{in} &= \hat{e}_{\mathbf{k},\sigma} E_0 e^{i\mathbf{k}\cdot\mathbf{r}} \\ \mathbf{H}_{in} &= \left(\hat{k} \times \hat{e}_{\mathbf{k},\sigma}\right) \frac{E_0}{\eta_0} e^{i\mathbf{k}\cdot\mathbf{r}}. \end{aligned} \tag{3.29}$$

Because coupling from all angles will be considered, there is a slight ambiguity in the definition of polarization used. Any consistent definition will have some strange consequences, but for the purposes of this analysis, a “TM-polarized” plane wave is one whose polarization vector has no $\hat{x}$-component and points in the $+\hat{y}$-direction, while a “TE-polarized” plane wave is whichever vector remains that is orthogonal to both the $\hat{k}$ and $\hat{e}_{TM}$ unit vectors.

Whatever the form of the impinging wave, it can be further reduced to components parallel to the $z = 0$ interface—the s-polarization—and those perpendicular to the interface—the p-polarization. Now, for a plane wave impinging on a uniform dielectric interface, the transmitted and reflected fields are given by the well-known results of phase matching, [12] and the transmitted fields can be written in terms of the incident ones as

$$\begin{aligned}\mathbf{E}_t &= \frac{2k_{1z}}{k_{1z}+k_{2z}}\mathbf{E}_i \\ \mathbf{H}_t &= \frac{2k_{1z}}{k_{1z}+k_{2z}}\frac{n_2}{n_1}\mathbf{H}_i\end{aligned} \tag{3.30}$$

for s-polarized light and

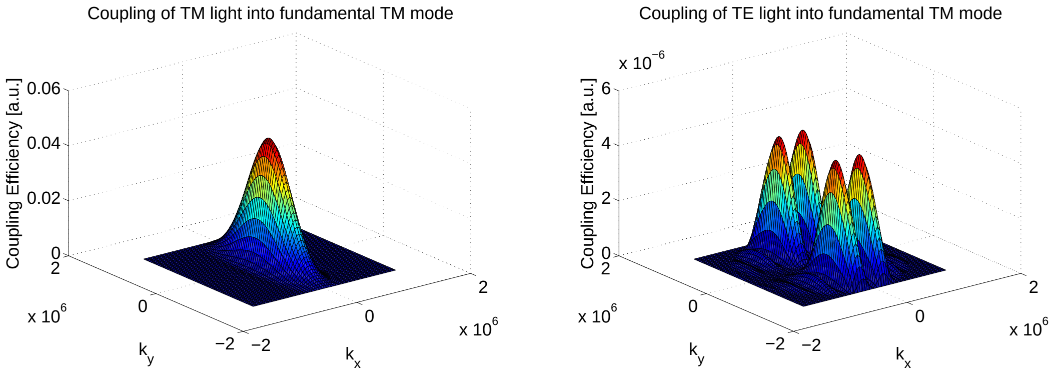
$$\begin{aligned}\mathbf{E}_t &= \frac{2k_{1z}\epsilon_2}{\epsilon_2 k_{1z}+\epsilon_1 k_{2z}}\frac{n_1}{n_2}\mathbf{E}_i \\ \mathbf{H}_t &= \frac{2k_{1z}\epsilon_2}{\epsilon_2 k_{1z}+\epsilon_1 k_{2z}}\mathbf{H}_i\end{aligned} \tag{3.31}$$

for p-polarized light. Since phase-matching holds, both k_{1z} and k_{2z} can be expressed in terms of the incident field's k_{1x} and k_{1y} as $k_{1/2z} = \pm\sqrt{(\frac{2\pi}{\lambda}n_{1/2})^2 - (k_{1x}^2 + k_{1y}^2)}$. The situation for coupling into a waveguide is not completely identical to that of a uniform dielectric interface, but since the incident field has been taken to be a pure plane wave and the reflection/transmission process is completely local, the transmitted field can be calculated by allowing k_{2z}, ϵ_2, and n_2 to be position-dependent and using the above relations.

Once the transmitted fields have been determined, finding the modal excitation is just a matter of exploiting the waveguide orthonormality relation expressed in (3.25). Assuming that the modes form a complete set of basis states for fields in the waveguide, the coefficient of the component of the transmitted field along mode β is

$$c_\beta = \hat{z}\cdot\frac{1}{4}\int\left(\mathbf{E}_\beta^*\times\mathbf{H}_t + \mathbf{E}_t\times\mathbf{H}_\beta^*\right)\, d^2\mathbf{r}_t, \tag{3.32}$$

and the power flowing in mode β is just $|c_\beta|^2$. If the power flow associated with the plane wave is held constant, the resulting quantity is then proportional to coupling efficiency. This procedure is done for as many values of k_{1x} and k_{1y} that are needed in order to get some desired angular resolution, and a profile of coupling efficiency versus angle is generated.



(a) (b)

Figure 3-4: Coupling efficiency into fundamental TM mode of 20 μm ridge structure at λ= 5 μm (a) TM light into TM mode (b) TE light into TM mode

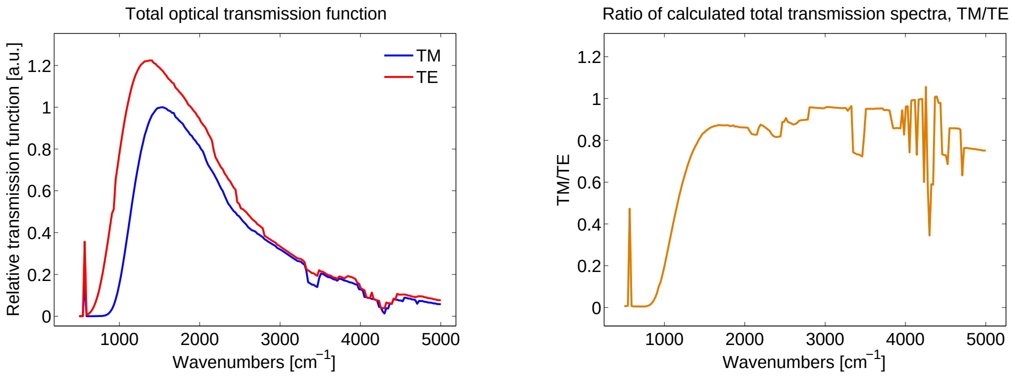
Figure 3-4 shows the computed coupling efficiency of light into the fundamental TM mode of a 20 μm ridge. Notice that light impinged in the TM polarization generally has the same shape as the Fourier transform of the mode profile (wide in x, narrow in y), even though it is quantitatively a little different. Also, while the TE coupling efficiency is non-zero, it is approximately four orders of magnitude weaker than the TM efficiency: as expected, light of one polarization cannot efficiently couple to waveguide modes of the other polarization. The only reason the TE efficiency is nonzero at all is that at high angles, the definition used for plane wave polarization breaks down and mixes the two components. (Regardless, on the $\hat{x}$- and $\hat{y}$-axis, the definition agrees with intuition and exactly zero efficiency is observed.)

Once the in-coupling efficiency has been found, the out-coupling efficiency is easy to obtain. Because the prior calculation has been done for all waveguide and spatial modes, a large part of the scattering matrix describing the system has been calculated. If each waveguide and spatial mode is represented by a single "port," then in calculating the coefficients c_β for each plane wave, the matrix entries $S_{\mathbf{k}\sigma,\beta}$ have actually been found. Furthermore, since the interface is a lossless, time-reversible structure, its scattering matrix is known to be symmetric [22], and the element $S_{\beta,\mathbf{k}\sigma}$—the quantity whose magnitude squared is the out-coupling efficiency—is precisely equal to $S_{\mathbf{k}\sigma,\beta}$.

There is one caveat associated with this analysis: because the impinged plane waves extend over all space and have infinite power flow associated with them, and because the free-space modes form a continuum of states, no precise determination can be made of the scattering matrix's proper normalization. Indeed, to obtain a physically meaningful result, some wave packet must be constructed that has finite spatial extent and finite power associated with it. For example, Figure 3-4 was calculated by assuming that the power associated with the field is just its intensity times the facet area (i.e., a sinc-like apodization function was used). To model the actual behavior of the optical system under consideration, it is assumed that light impinges from the high-NA objectives isotropically, and that light from every possible **k** that could from the objectives does so with equal weight. For the reflective objectives that were used, light can only couple from angles between 18° and 35.5°, and the apodization function is essentially a "donut" in **k**-space.

Regardless, the main quantities of interest in absorption measurements do not require knowledge of the absolute coupling efficiency, only the relative contribution between TE- and TM-polarized light. To complete the calculation, an equal amount of light in each polarization is first impinged with the aforementioned donut profile, and the resulting power in each mode is calculated. Then, the modes are propagated down the waveguide using their calculated losses, their out-coupling efficiency is calculated using the same donut apodization, and the relative power in each polarization is compared. In addition, the average Γ and n_{eff} are calculated by weighting their values for each mode by the final amount of power in said mode. All of this is performed at a broad range of wavelengths, ranging from 2 μm to 20 μm.

Figure 3-5(a) shows the simulated transmission functions for TM-polarized light and for TE-polarized light, as well as their computed ratio. At very low frequencies, practically no guided modes exist in the active region, as the waveguide is in cutoff. As the frequency is increased, the fundamental lateral mode comes into existence, and an optimum value is reached in which the output angle of the fundamental mode



(a) (b)

Figure 3-5: Total transmission function of TM- and TE-polarized light through objectives and 20 μm ridge (a) Both transmission functions (b) Ratio of TM transmission to TE transmission

best matches that of the objective. At still higher frequencies, higher-order lateral modes become supported, but since none couple very well to the spatial modes of the objective, both polarizations eventually roll off. More importantly, the ratio between the polarizations may start near zero, but quickly levels off to about 0.85 past a frequency of 1500 cm^{-1}. This is crucial for absorption measurements, which require that the TM and TE modes have similar properties (or that the difference can be easily quantified). However, at higher frequencies, beginning around 3000 cm^{-1}, while the ratio generally stays between 0.8 and 1.0, it becomes "noisy." In fact, this noisiness is not a discretization artifact or even a simulation shortcoming: it is a direct consequence of the lateral modes of the waveguide. At high frequencies, dozens of modes can be supported, and their nontrivial couplings begin to superimpose, creating a chaotic pattern that is all but impossible to exactly describe. It will be soon evident that this effect is real and actually does appear in the measured data.

Figure 3-6 shows the average Γ and average n_{eff} computed as a function of frequency. Note that once again, both quantities become noisier at high frequencies, thanks to the influence of multiple modes. Unexpectedly, n_{eff} is held relatively fixed by the fact that all III-V materials have a refractive index near 3. The confinement factor is a little more sensitive: as frequency is increased, Γ increases until the modes are pulled into the active region, eventually levels off, and finally decreases once the

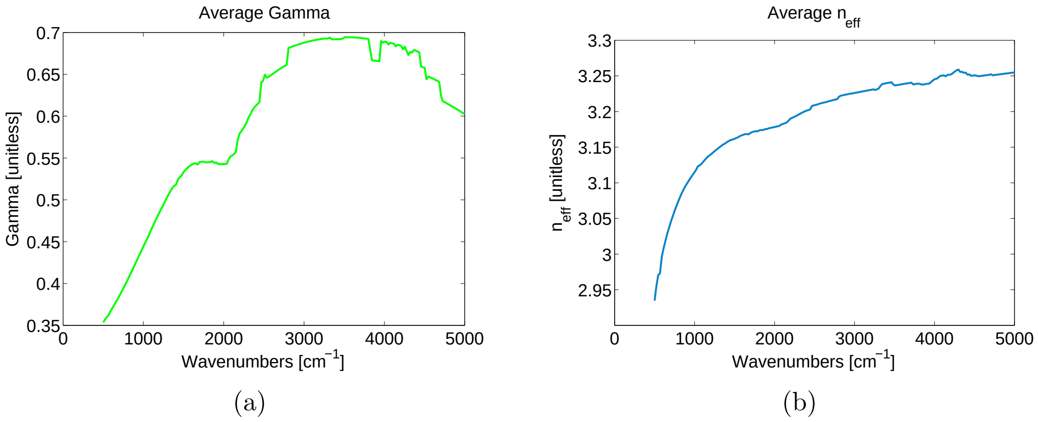


Figure 3-6: Average $\Gamma(\nu)$ and average $n_{\text{eff}}(\nu)$ determined by weighting modal values with coupling factors (a) Average $\Gamma(\nu)$ (b) Average $n_{\text{eff}}(\nu)$

modes begin to be pulled into the waveguide layers. Though neither changes very quickly over the primary range of interest (above 1500 cm^{-1}), because the variation is nontrivial and each quantity factors directly into absorption, the dependence needs to be included.

Chapter 4

Experimental setup

Several types of measurements were performed, including the broadband absorption measurement, the Fourier-Hakki-Paoli waveguide loss measurement, laser emission measurements, coupling measurements, and photovoltaic measurements. This chapter describes the methodology behind them, in addition to practical concerns in their implementation.

4.1 General considerations

4.1.1 Device mounting

After cleaving to a desired length, all devices were mounted with indium solder on a copper mount that served as a heat sink for biased devices. Gold wires were then bonded from the device under test to a gold contact pad, also affixed to the mount. For spectroscopy applications, it was important that the contact pad not be placed directly behind or in front of either laser facet, to ensure that it did not block light entering or exiting the cavity. Placement of the device on the mount was another critical concern, because each facet that was not flush with an edge of the mount would have its light blocked by the mount itself. Not only would this halve the signal-to-noise ratio, but it would also introduce periodic “ringing” in the coupling efficiency, by virtue of the fact that a relatively frequency-independent coupling ratio requires that the input beam matches the waveguide modes as best as possible. Generally, two types of mounts were used, pictured in Figure 4-1. The flat mounts had one device facet flush with the edge of the copper and one facet located over the flat part of the mount. By contrast, the upright mounts had both facets flush with the copper edge. Owing to their superior transmission, upright mounts were used in most cases, except when devices had already been mounted prior to transmission spectroscopy and could not be removed with destroying the device.

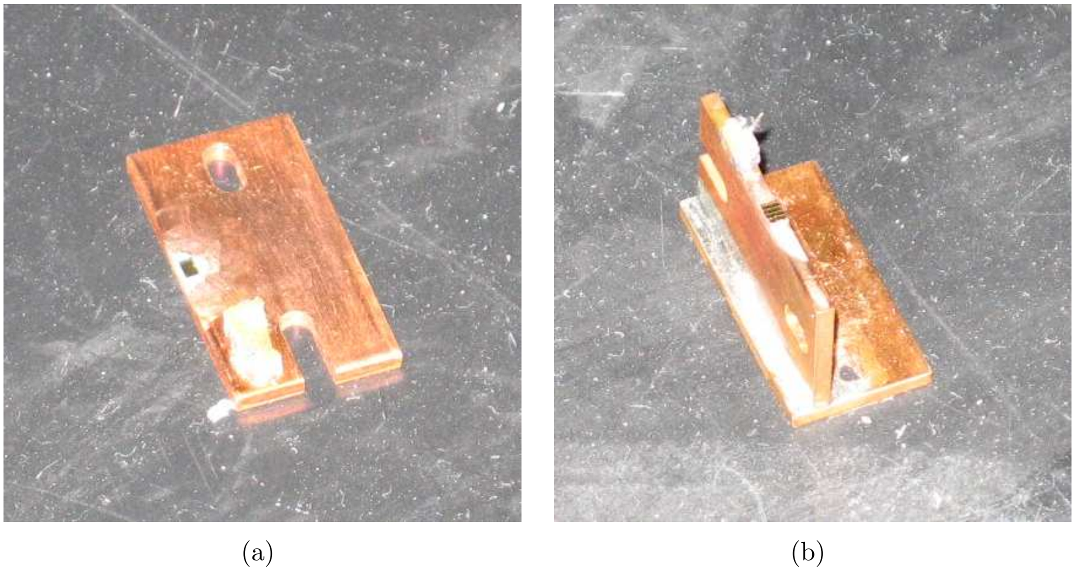

(a) (b)

Figure 4-1: Device mounts used for transmission experiments (a) Flat mount (b) Upright mount.

4.1.2 Fourier Transform Infrared Spectrometry

At the heart of any type of spectral measurement in the long wavelength regime is the Fourier Transform Infrared Spectrometer (FTIR). Unlike other types of spectrometers, which separate broadband light into its spectral components and measure the power associated with each to generate a spectrum, the FTIR samples an interferogram, a scaled version of the Fourier transform of the spectrum and an analog of the time-domain power signal. The basic schematic of an FTIR is shown in Figure 4-2.

Essentially, the FTIR is just a Michelson interferometer with a moving mirror controlled by a computer. As the mirror scans through its range of motion, the light

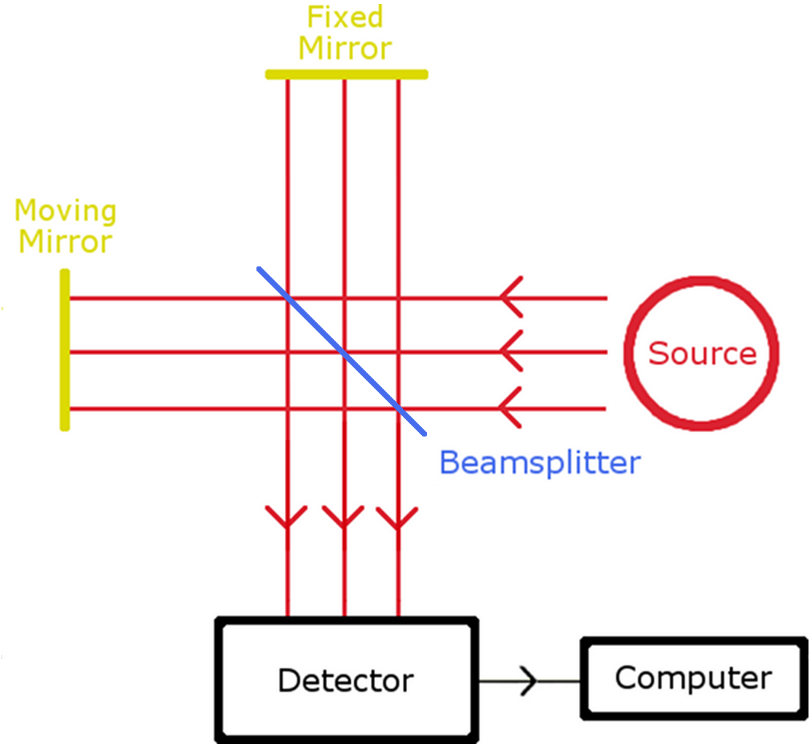


Figure 4-2: FTIR Schematic. Light enters the beamsplitter, is sent down two paths with different lengths, and is recombined at the detector.

that has traveled down one arm sees a different optical path than the other, and an interference pattern is produced that is then recorded by a computer. Analytically speaking, the intensity at the detector can be shown [23] to be a scaled, shifted version of the Fourier cosine transform of the optical spectrum:

$$\begin{aligned} I(\delta) &= C_1 + C_2 \int_0^{\infty} |F(\sigma)|^2 \, cos(2\pi\sigma\delta) d\sigma \\ |F(\sigma)|^2 &= C_3 + C_4 \int_0^{\infty} I(\delta) cos(2\pi\sigma\delta) d\delta, \end{aligned} \tag{4.1}$$

where C_1, C_2, C_3, and C_4 are some constants, $I(\delta)$ is the intensity recorded at an optical path difference of δ, and $|F(\sigma)|^2$ is the spectral intensity at the wavenumber $\sigma = 1/\lambda$. After a sufficiently large range of positions has been recorded, the interferogram $I(\delta)$ is generated, and the optical spectrum can be found through (4.1). It should be emphasized that Figure 4-2 is just a cartoon: because the interferometer is a linear system, the detector and source can be placed anywhere in the system, provided that the light passes through the interferometer. In other words, for spectroscopy applications, an absorptive element's location is irrelevant.

The FTIR used for these measurements was a Nicolet 8700 spectrometer manufactured by Thermo Fisher Scientific Inc. It has a beamsplitter made of potassium bromide that is completely transparent to mid-IR, as well as gold interferometer mirrors that serve as near-ideal reflectors. The FTIR contains an internal reference helium-neon laser that is used to monitor mirror position, and is able to resolve position with a resolution as low as $\frac{1}{4}\lambda_{\mathrm{HeNe}} = 158$ nm, allowing it to detect wavelengths as low as 316 nm. The mirror can travel a distance of approximately 4 cm from the zero-path-difference position, giving it a spectral resolution down to 0.125 cm^{-1}. In addition, it contains an internal source made of a thermally-stabilized silicon-carbide element that produces blackbody radiation in the mid-IR when heated. Though the source is broadband and emits a fairly high power of 14 mW, its brightness is limited by the fact that it is also quite large (a few millimeters across). It can also accept light from an external source in lieu of the internal one.

The FTIR can use any number of detectors, but the primary ones used in these experiments were those made of mercury cadmium telluride (MCT). MCT is a narrow bandgap material that can respond to light in the range of 600 cm^{-1} to 11700 cm^{-1}, and with a voltage responsivity of 7000 V/W and a detectivity of $D^* = 4.7 \cdot 10^{10} \mathrm{cm}\,\mathrm{Hz}^{1/2}\,\mathrm{W}^{-1}$, it stands as the most sensitive detector that can cover the whole mid-IR. In addition, its response time is short, and it is capable of detecting light that is modulated at 175 kHz. The main difficulties associated with MCT are the fact that it must be cooled by liquid nitrogen to maximize its sensitivity, and the fact that the detection element is only 250 μm by 250 μm wide, in order to minimize its noise figure.

4.2 Absorptivity measurement

Figure 4-3 shows the setup used for QCL absorption measurements. First, light that has been emitted by the internal FTIR source passes through the interferometer and

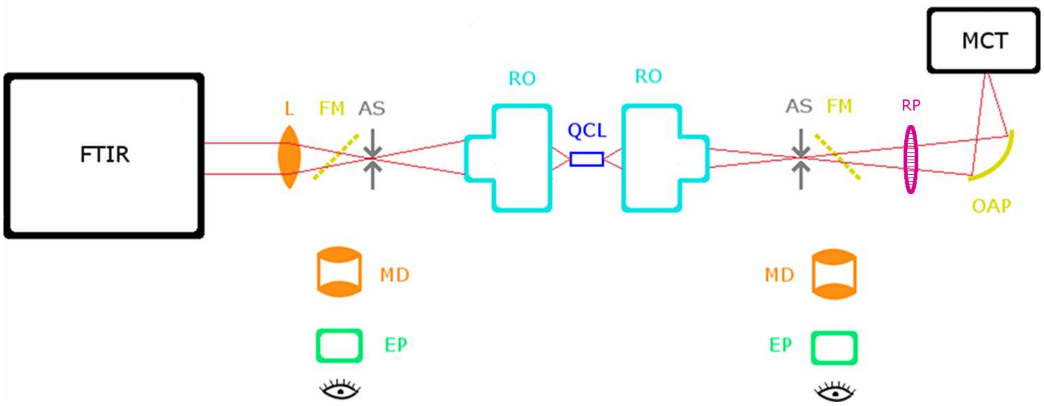
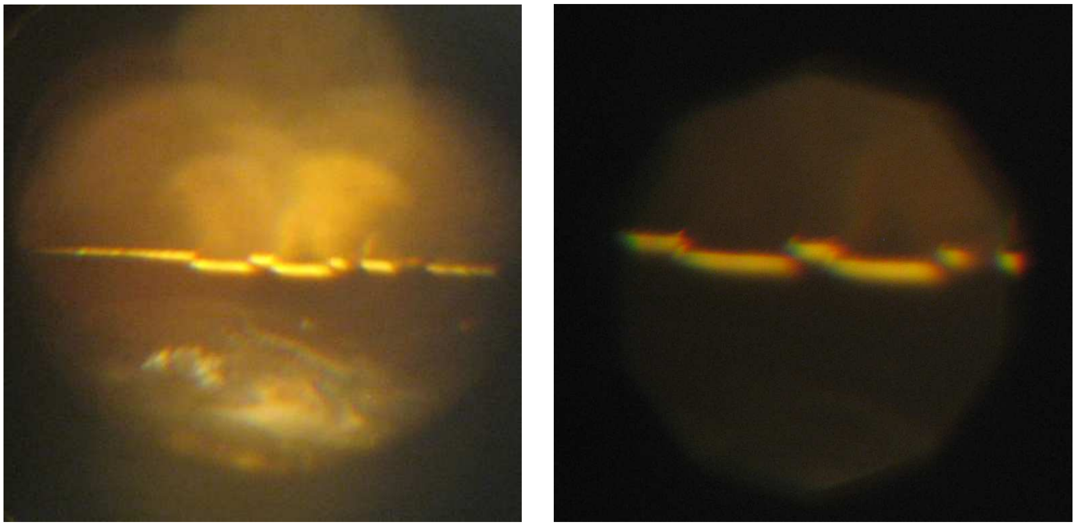


Figure 4-3: Absorption measurement schematic (not to scale). Light passes from left to right, assuming all flip mirrors are in the down position.

out its exit port, where it is collimated. It then passes through an f/1 AR-coated zinc selenide lens with a focal length of 2”, which focuses the light onto an adjustable aperture and into a 15x reflective microscope objective with a numerical aperture of 0.58. When everything has been properly aligned, the aperture rests in an image plane of the objective, and selects only that light which would fall on the front facet. Light passes through the waveguide, is collected by another objective that refocuses the light to a different image plane, and a second aperture selects only light that has emanated from the rear facet of the laser. Lastly, light passes through a rotatable wire-grid polarizer, and is collected by an off-axis parabolic mirror that focuses the beam onto an MCT detector’s element.

The importance of careful alignment of the laser and of the optics cannot be understated. While considerable effort was expended creating optical systems in which alignment would be accomplished by maximizing some chopped signal through the adjustment of some translation stages, ultimately, the objectives used are too fast for such an attempt and the search space is too large. Even when some seemingly optimal configuration was found, there was no guarantee that the configuration was not some type of local minimum. To make matters worse, there was never any guarantee that light was actually passing through the QCL waveguide, as opposed to through the substrate or over the top of the device completely. Thus, no measurements could be made with any confidence as to their validity.

To rectify this, microscope objectives were used that are purely reflective and can operate from the infrared to the visible without any sort of frequency-dependence. With these, ideal alignment can be accomplished by simply looking at the image of the device through each objective and adjusting accordingly. For practical reasons, the image of each aperture was relayed several inches away from their physical location via a pair of gold flip mirrors and a matched achromatic doublet, and each image was then magnified ten times by a microscope eyepiece, creating a total magnification of 150x. With proper illumination, both the aperture and the objective's image plane are visible in the eyepiece, and alignment becomes a simple matter of moving the laser facet into the center of the aperture. Figure 4-4 shows pictures of a device facet as seen through the eyepiece, both with the aperture fully open and with the aperture partially closed. Note that the gold contact layers are clearly visible, as is the facet.



(a) (b)

Figure 4-4: Pictures of 20 μm ridge facet as seen through objective. (a) Aperture completely open (b) Aperture partially closed, zoomed in.

Once the device facets have been aligned to the center of each aperture, the apertures are closed to leave only the facets left unblocked, and the mirrors used for image relay are flipped down to allow light from the FTIR to pass through the system. Finally, the ZnSe lens and off-axis parabolic mirror are positioned. Unlike the objectives, whose proper positions are very difficult to find without visible cues, the lens and mirror positions are not so crucial, and can be found by placing a mechanical chopper in the beam path, measuring the signal detected on the MCT with a lock-in amplifier, and optimizing for increased throughput.

To actually perform the absorptivity measurement, the rotatable polarizer is placed in the beam path, and is oriented in either the TE or TM polarization. Though the light emanating from the second objective is not strictly collimated, it diverges slowly enough that polarization mixing does not occur. Indeed, this was verified by biasing a laser placed in the system (known to be pure TM), inserting the polarizer, and comparing the ratio of each polarization measured on the detector. The measured extinction ratio was better than 200:1, and is within the range quoted by the manufacturers.

However, before absorption measurements are taken, there are several metrics that should be measured to characterize the optical system. First and foremost is the total amount of light that couples through the device and is collected by the detector when no polarizer is present. When the beam was modulated by a mechanical chopper at 270 Hz and light was passed through a 20 μm ridge device, a signal of about 20 mV was typically measured at the detector with a lock-in amplifier, corresponding to a power flux of about 22 nW. (The actual amount depends on device geometry and mounting, the primary factors being the width of the device and whether the device is mounted flat or upright.) Power fluxes as high as 45 nW have been observed for well-mounted broad-area devices, and values as low as 500 pW have been observed for narrow ridge structures. Slightly damaged structures can completely extinguish transmission.

While power levels of 20 nW are sufficient to run a scan with a resolution of 16 cm^{-1} and a signal-to-noise ratio better than 100:1 in just a few minutes, some considerations will force the acquisition time much higher. The insertion of the polarizer must by definition cut the signal approximately in half, but since TM light is strongly absorbed by the QCL, it can cut total TM transmission by as much as a factor of ten. For signal levels of that magnitude, the scan time may have to be increased to an hour or more to obtain the requisite signal-to-noise ratio. Unfortunately, there is little that can be done to improve this. It is a well-known result from optics that any system has a fundamental limit imposed on its ability to concentrate light, from geometry *alone*. [24] More precisely, if an isotropic light emitter—including glowing infrared sources—has an area of A and emits uniformly into a solid angle Ω, the quantity $A\Omega$ (known as the etendue) is at best invariant, and is at worst made larger. In other words, the only way to focus the light of the glow-bar to a smaller area is to use a system with a larger focusing angle, that is, with faster optics. At the same time, however, the maximum coupling angle of the QCL is fixed by the critical angle of the waveguide, and the area is fixed by the spatial extent of the optical modes. Therefore, the maximum power flux is effectively set by the brightness of the source, provided that the etendue has not been decreased through the use of overzealous apertures or aberrational optics. The only way to increase the brightness of a glowing source is to increase its temperature, but this cannot be done for practical reasons, as the silicon-carbide element would vaporize. Other types of sources could also be used, including halogen lights and mid-IR LEDs, but the former is opaque to mid-IR radiation and the latter is too narrowband.

The optical system should also be characterized with the device removed, in order to remove any unwanted effects that it might introduce. Without any polarizer, approximately 1 μW of power transmits through the system; with the polarizer, approximately 400 nW transmits. However, the system is not completely isotropic, and more power transmits in the TE polarization, especially at high frequencies. By carefully measuring the transmission spectrum of each direction and calculating their ratio without changing the apertures' shape, the anisotropy could be extracted from all absorption measurements. Figure 4-5 shows the TE and TM transmissions as a function of wavenumber. Of note is the presence of the atmospheric absorption lines on each spectra, principally due to atmospheric water and carbon dioxide absorption. Because the system was too large to purge with nitrogen, these features appeared on every spectrum. Even so, as they were not polarization-dependent, they could not bias the measurement in any way, except to add a small amount of noise.

Once the transmission spectrum of the system with and without a device has

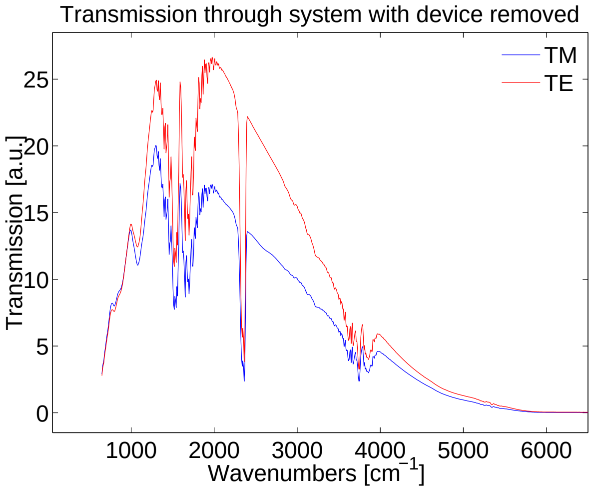


Figure 4-5: Relative transmission of optical system without device.

been measured in each polarization, calculating the absorptivity is just a matter of exploiting the fact that the TE light cannot see intersubband transitions. Including normalization, the absorptivity can be written down as

$$\alpha(\sigma) = -\frac{1}{L} \ln \left(\frac{\frac{P_{\mathrm{TM}}(\sigma)}{N_{\mathrm{TM}}(\sigma)}}{\frac{P_{\mathrm{TE}}(\sigma)}{N_{\mathrm{TE}}(\sigma)}} \right), \tag{4.2}$$

where L is the device length, $P_{\mathrm{TM}}(\sigma)$ and $P_{\mathrm{TE}}(\sigma)$ are the spectra measured with the device present, and $N_{\mathrm{TM}}(\sigma)$ and $N_{\mathrm{TE}}(\sigma)$ are the normalization spectra measured without the device present.

Since none of the devices tested could survive room-temperature CW operation, biased measurements could only be performed by pulsing the laser with a sufficiently low duty cycle. Spectra could then theoretically be measured by modulating the input beam with a slow mechanical chopper, modulating the laser at a high frequency, and using a pair of lock-in amplifiers to detect only that light which has been modulated by both, at the sum and difference frequencies. In practice, for ridge devices pulsed with a 5% duty cycle and modulated at 10 kHz, when the input beam was modulated at 270 Hz, a signal-to-noise ratio of 30 dBV was observed at the sum and difference frequencies, and would have been sufficient for spectral measurements. However, the FTIR used to measure spectra had experienced technical difficulties that caused its step-scan mode to malfunction, and as a result, measurements of this type could not be taken. Therefore, absorption measurements were restricted to flat-band conditions.

4.2.1 Coupling efficiency

Even though the normalization has removed essentially all of the spectral dependence of the system's optics, it is important to note that the one effect it has not been able to remove is that of coupling differences between the polarizations. As already discussed, simulation suggests that polarization differences level off to about 0.85 at high frequencies, creating a shift of a couple cm^{-1} on the final absorption spectrum. The absolute coupling efficiencies cannot be determined for both polarizations (as intersubband transitions will always interfere with the measurement of TM efficiency), but by ensuring that neither aperture size is changed between the TE normalization measurement and the TE device measurement, a total transmission efficiency can be assigned. Figure 4-6 shows the measured efficiency of two devices which came from the same wafer and were processed the same way, but which were cleaved to different lengths. It is important to note that past the waveguide's cutoff frequency of about 1100 cm^{-1}, transmission levels off to an approximately constant value. While it is tempting to attribute the variation of the efficiency in the transmission range to a

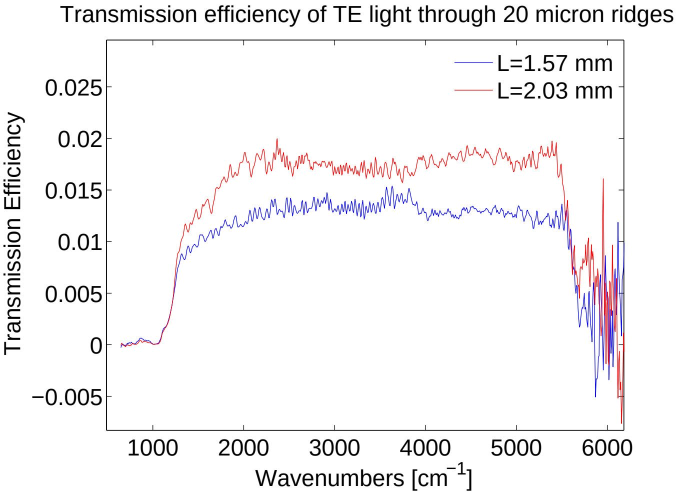


Figure 4-6: Transmission efficiency of TE light through devices from same wafer but cleaved to different lengths.

poor signal-to-noise ratio, the variation magnitude falls several standard deviations above the detector noise fluctuation. In fact, the variation actually comes from the seemingly chaotic effect of the multiple modes of the waveguide. Past 5000 cm^{-1}, the polarizer begins to lose transmissivity (see Figure 4-5), and the noise level increases greatly. Nonetheless, an important feature that appears on all device spectra can be seen through the noise: a sharp drop-off in transmission past 5600 cm^{-1}. This effect is attributed to the InGaAs waveguide layers, whose band gap is about 0.7 eV and would isotropically absorb all radiation whose frequency is higher than this.

The main disadvantage of measuring coupling efficiency this way is that, once again, the efficiency definition depends on the input aperture shape. For example, if the aperture were opened fully, the apparent efficiency would seem to drop, as the input beam's focus would become large and would contain a lot of power, but the exact same amount of light would couple into the waveguide. In addition, the apparent efficiency is dependent on the loss of the waveguide modes, as well as device length. As a more absolute measurement, it is useful to change the existing setup by replacing the FTIR with a flat mirror, as shown in Figure 4-7. A laser is then biased past threshold, and the amount of light received by the detector is measured by a lock-in-amplifier. When the flat mirror is aligned normally to the collimated beam, light that escapes the cavity from the front facet can be reflected back into it, enhancing the effective reflectivity of the front facet and increasing the amount of light seen by the detector. The power at the detector can then be compared to the power at the detector when the front facet is blocked, and the result can be used as an estimate for the product of the coupling efficiency. This measurement was performed,

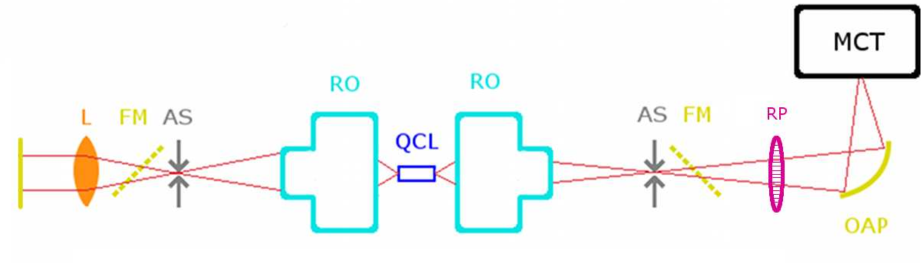


Figure 4-7: Setup used to measure coupling efficiency.

and it was found that for 20 μm ridges, approximately 8% of the emitted light would couple back into the waveguide. As this estimate does not include a factor of the waveguide loss, it is much more representative of the true efficiency.

4.3 Fourier-Hakki-Paoli loss measurements

Not only can the intersubband properties of waveguides be probed by this method, but so can their waveguide losses. When a pulse of light enters a lossy waveguide from the front facet, it propagates down the waveguide to the other end, where a fraction of the power escapes and a fraction reflects. The power that reflects will propagate back to the first facet, partially reflecting and partially transmitting, and the reflected power repeats the process. In the time domain, an observer at the back facet would observe the initial pulse, a secondary pulse a time $2L/c$ later, a third pulse after another time $2L/c$, etc. The amount of power contained within each pulse would decrease geometrically by a factor related to the mirror losses and to the waveguide losses.

In the frequency domain, this gives rise to the well-known Fabry-Perot fringes, and one can use the technique pioneered by Hakki and Paoli to find the waveguide losses, in which the ratio between fringe heights is exploited. [4] However, when dealing with low-finesse cavities, if one tries to find the losses from the spectrum, one also finds that the calculation is extremely sensitive to noise. Fortunately, FTIR provides a natural alternative, since it takes data in the time domain, not the frequency domain. By measuring a transmission spectrum at high resolution, one can simply look at the interferogram generated by the FTIR to measure the average loss over the whole frequency span. (Alternatively, one could window the spectrum in some region and Fourier transform to measure the average loss only in that region.) As with the coupling measurement, it is essential that this method only be used with TE light, since TM light is affected by intersubband transitions. Figure 4-8 shows a logarithmic plot of the interferogram associated with the TE transmission of a 20 μm Razeghi-599 device. The zeroth order peak represents light that escaped the waveguide after the

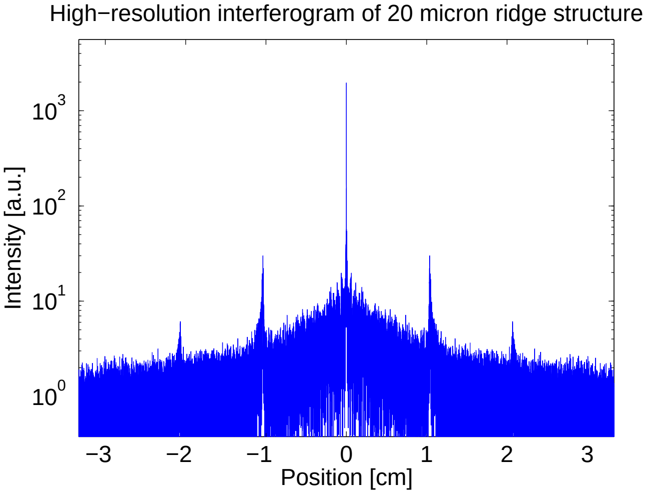


Figure 4-8: Fourier transform of TE transmission spectrum of 20 μm ridge.

initial pass, the first order peaks represent light that reflected once, and the second order peaks represent light that reflected twice. Third order reflections fell below the detector noise level and could not be measured. Note that the distance between each successive peak gives a precise measurement of the phase delay associated with two passes of the waveguide, and is an accurate "ruler" for the length of the device.

To actually extract the loss, one makes use of the ratio of the heights of the higher-order harmonics, which are related by a simple factor of [5]

$$\frac{p_{n+1}}{p_n} = Re^{-\alpha L}, \tag{4.3}$$

where R is the mirror *power* reflectivity ($R \approx 0.3$) and α is the *power* loss (including both waveguide and potentially intersubband loss). Strictly speaking, n must be also be nonzero, since the zeroth-order peak contains an extra delta function corresponding to the average power of the spectrum and cannot be used. In all devices tested, at least two harmonics were visible above the noise floor, permitting the use of this type of measurement. However, because high-resolution spectra require significantly longer acquisition times than low-resolution ones and because the heights of peaks decayed so quickly, third harmonics were observed very rarely. Still, this measurement proved to be a useful diagnostic for problems with waveguides that had non-obvious problems, such as scattering defects or spurious modes. In the former case, a slight crack appeared in the waveguide a third of the way across the device, a problem that revealed itself through the presence of many harmonics that were spaced unevenly. In the latter case, two sets of harmonics were generated with different mode spacings, suggesting that guided modes existed with measurably different phase velocities.

Of course, Fourier-Hakki-Paoli is not just limited to actual transmission experiments. Because Fabry-Perot fringes also appear in the emission spectra of lasers near threshold, the same method can be applied to find the properties of lasers using light generated *internally.* In fact, because (4.3) actually applies in a time-dependent sense, and not just at the peak of each harmonic, it can be theoretically used to find the frequency-dependent absorption. For broadband transmission experiments, this is impractical because the required temporal-spatial extent of the interferogram is too large, and a small amount of noise is enough to completely overwhelm the measurement. However, the narrowband emission of a laser biased below threshold is mathematically well-behaved, and its Fourier transform is a smooth function.

The major limiting factor is whether enough light can be collected to perform the measurement in an appreciable amount of time, so a very strong signal is required. Because only near-threshold CW operation could provide enough light, and the lasers tested could not survive room-temperature CW operation, all devices measured in this way were cooled in a liquid nitrogen cryostat and biased just below threshold. The output of the devices was collimated by a fast off-axis parabolic mirror, passed into the input port of the FTIR, and collected by an MCT detector. As the FTIR has no way of distinguishing between the thermal radiation generated by room-temperature blackbody sources and the signal emitted from the laser in this configuration, a background scan must first be run with the device turned off, with the result being subtracted from all subsequent spectra. In order to get an idea of the bias-dependence of the absorption curves, spectra are run for many different values near threshold. Figure 4-9 shows an example series of interferograms (from measurements performed by Hsu [3]) for one of the Razeghi-605 series of devices, plotted on a linear scale. The most jarring differences between this and the transmission experiment are the greatly increased signal-to-noise ratios and the noticeably smaller rate

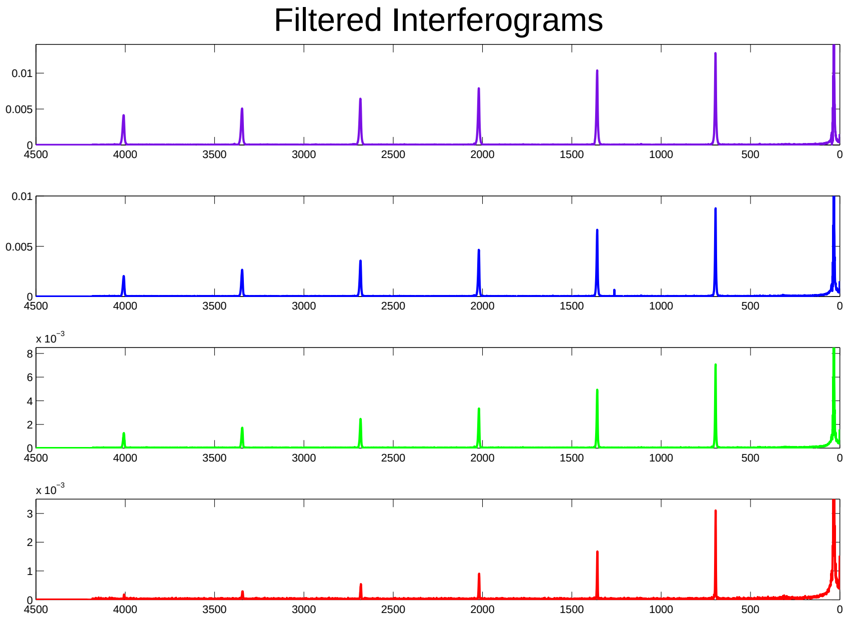


Figure 4-9: One-sided interferograms of nitrogen-cooled device at different biases. The topmost interferogram is at the highest bias, while the bottommost is at the lowest bias.

at which the harmonics decay. In fact, as many as ten peaks have been observed on the interferograms of shorter devices, which do not hit the limits of the resolution of the FTIR as this device did. Why this should happen is intuitively obvious: a laser biased near threshold sees almost no net loss, and thus the peaks should decay very slowly. On the other hand, losses in the the transmission experiment are often of the order of tens of cm^{-1}, and so it is no surprise that the peaks should decay very quickly.

Once the cleaned-up interferogram has been generated, each of the peaks are isolated, and the ratio that fits them best is found. Because the fitting was performed in a least-squares sense, the inclusion of noisy high-order harmonics does not interfere with the cleanliness of the ratio obtained; to the contrary, including those harmonics drastically improves the signal-to-noise ratio of the final result. Finally, the ratio is inverse Fourier transformed, and the frequency-dependent losses are extracted through an application of (4.3).

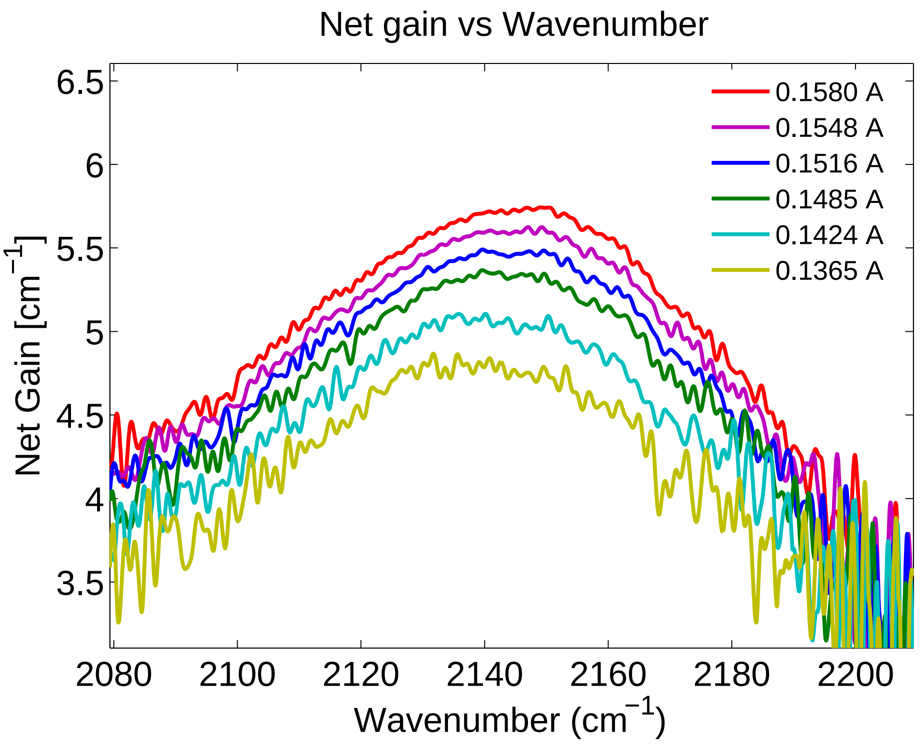


Figure 4-10: Gain spectrum of device measured at different biases.

Figure 4-10 shows the results of this analysis (performed by the author). Note that instead of plotting the losses, the gain has actually been plotted. By definition, the waveguide of a device biased near threshold must be providing as much gain as is necessary to compensate for the mirror losses, and so it is no surprise that the total “losses” are negative. Also evident is the presence of a shift in the peak of each gain curve as the applied bias is increased, a classic effect from quantum mechanics. Going one step further and plotting the peak value of each curve as a function of bias, one finds that the peak value obeys a linear relationship with current, from which one can extract the current back to zero and estimate the waveguide losses.

Of course, this method is not without its flaws. Because it relies on the internal emission of the laser, it is intrinsically narrowband, and will always be limited to bandwidths of about 100 cm^{-1}. Furthermore, it requires that the device under test be fairly robust, and is not conducive to devices that operate poorly or quickly fail under CW operation. Lastly, there is no information that can be obtained from the method that cannot theoretically be obtained from a transmission experiment.

4.4 Photovoltaic measurement

One technique that can be used to obtain some of the same information as the absorption measurement but with a much simpler setup is a photovoltaic spectral measurement. Essentially, the front half of the absorption setup is used, but instead of using an MCT detector to measure the spectrum of light transmitted through the QCL, the QCL is itself used as a detector. Light that enters the waveguide induces upward transitions in the active region, and because QCL active regions are asymmetric, electrons in excited states will decay preferring one direction or the other. The collective effect induces a photocurrent in the device, or if the top and bottom contacts are electrically open, a photovoltage. Assuming that the photovoltage/current is a linear function of the power absorbed, the device can essentially be considered a photodetector, one whose responsivity depends on the band structure of the QCL active region. Therefore, when the FTIR takes a spectrum, the result is the internal source's spectrum times the responsivity of the QCL, and is correlated to the absorptivity of the device. Figure 4-11 contains the setup of a photovoltaic measurement. The QCL can either be connected directly to the FTIR in linear scan mode as shown or connected via a lock-in amplifier and used in step scan mode. When linear scans can be done, they are preferred because they require less acquisition time in the absence of background light. In addition, while various biases and modes were attempted, the most successful scans came from devices that were unbiased and used in pure photovoltaic mode. While it was indeed possible in some instances to pulse the device, chop the incident beam, and measure a small change in the current drawn

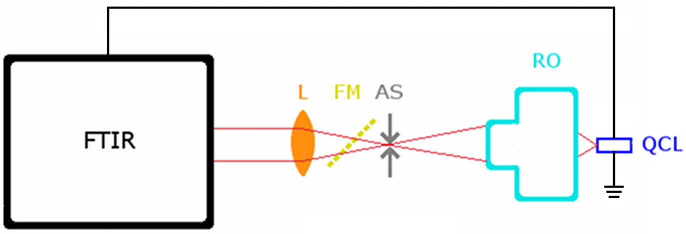


Figure 4-11: Photovoltaic spectral measurement.

by the device with a lock-in amplifier, the bias current would more often than not exceed the already substantial dynamic range of the amplifier.

The primary disadvantage of a PV measurement is that it requires detailed knowledge of the transport properties of a QCL under illumination to interpret. More specifically, one must know how the excited states will decay in order to estimate the rate at which electrons can transport across a module. For example, an electron in the second excited state of an isolated quantum well could do nothing but relax down to the well's ground state, producing no current. At the other extreme, the excited states of a bound-to-continuum QCL design can easily couple to the continuum, which at zero bias forms a "staircase" of states that permits transport across a module through the emission of LO-phonons. The injectorless QCL designs tend to fall somewhere in the middle, permitting some transport but not nearly as much as the bound-to-continuum designs. At the same time, even if one could reliably estimate the contribution of each excited state to the photocurrent, photovoltaic measurements are even more complex, requiring knowledge of how the electrons distribute themselves once they have reached the last module and form a space charge.

Of course, the PV measurement is not without its advantages. In addition to its relative simplicity, the mechanism through which the band structure is obtained is fundamentally different than the absorptivity measurement, and sometimes provided a better signal-to-noise ratio. Consider a large QCL, one that has been cleaved to be longer than 3 mm. (For some devices, long devices were the only way to achieve lasing operation.) In the absorptivity setup, accurate measurement of a strong absorption line requires that very small power levels be measured in the transmitted spectra. However, the amount of TM light that is collected is proportional to $e^{-\alpha L}$, and the exponential dependence ensures that this is all but impossible for long devices. In contrast, the photocurrent is proportional to the *energy* absorbed by the QCL, which is in turn directly proportional to the absorptivity. Put another way, the PV measurement is a direct measurement of absorption, while the transmission measurement requires a substantial amount of noise-sensitive post-processing to get its result. Even though the transport factor makes precise determination of the oscillator strengths impractical, it is nonetheless very useful for the determination of the transition energies of the system, and also provides confirmation of the veracity of the transmission experiment.

Chapter 5

Results and analysis

Finally, the results of the band structure simulations and measurements are compiled and analyzed. This chapter is divided into three sections, each of which contains the results from a single group of structures.

5.1 MIT-Razeghi devices

The MIT-Razeghi devices, which made up the majority of the devices that were tested, are mid-infrared QCLs grown and fabricated by MIT Lincoln Laboratory as part of the EMIL project. They were based on designs proposed by Evans and Razeghi in 2007, who had demonstrated the design's robustness and efficiency, producing 0.675 W of continuous-wave power at 4.8 μm with a wall-plug efficiency of 9.3%. [25] The structure is a four-well design fabricated with strain-balanced $\mathrm{Ga}_{0.331}\mathrm{In}_{.669}\mathrm{As}$ wells and $\mathrm{Al}_{0.645}\mathrm{In}_{.355}\mathrm{As}$ barriers, and when biased should have a high injection efficiency, low levels of tunneling into the continuum, and low interface scattering. [26]

Figure 5-1 shows the band structure and wavefunctions of the lowest-doped Razeghi structure under flat-band conditions. Though no direct correlation can be made with the structure's energy levels under zero bias and under design bias, generally speaking, the pseudo-continuum of low energy states become the design's injector, and the upper-lying states become the excited states of the laser. Notice that since electrons in excited states can couple efficiently to the injector, this design is particularly well-

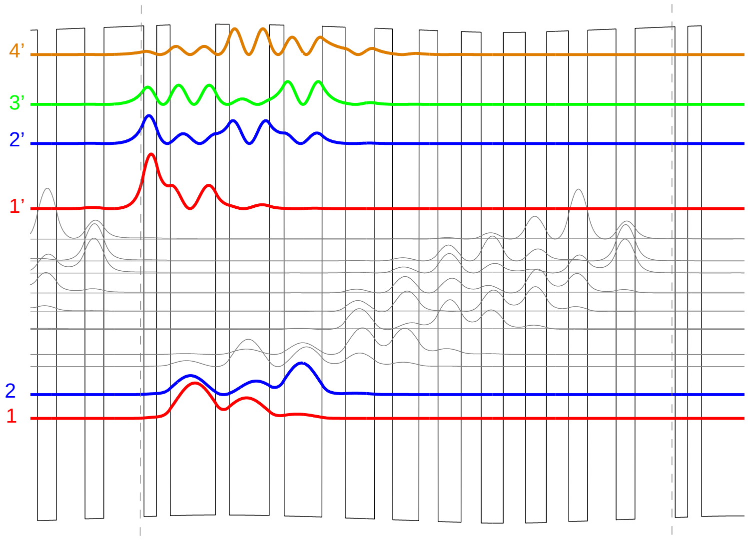


Figure 5-1: Band diagram of Razeghi-599 structure. Starting from the left-most dashed line, layer thicknesses are 12/13/43/13/38/14/36/22/28/17/25/18/-22/**19**/**21**/**21**/**20**/21/18/27/18/38 (in Angstroms, doped layers in bold). Wells are $\mathrm{Ga}_{.331}\mathrm{In}_{.669}\mathrm{As}$, barriers are $\mathrm{Al}_{.648}\mathrm{In}_{.355}\mathrm{As}$.

suited for photovoltaic measurements. In addition, despite the multitude of states in the system, only the lowest two are significantly populated at room temperature, so the absorption measured should therefore be relatively simple.

Many different types of MIT-Razeghi devices were grown and fabricated. Firstly, four different wafers were grown, each with an active region of 30 modules but with different dopings. The wafers are denoted by the names Razeghi-599, Razeghi-605, Razeghi-606, and Razeghi-627, and have respective nominal dopings of $1 \cdot 10^{17}\,\mathrm{cm}^{-3}$, $2 \cdot 10^{17}\,\mathrm{cm}^{-3}$, $3 \cdot 10^{17}\,\mathrm{cm}^{-3}$, and $4 \cdot 10^{17}\,\mathrm{cm}^{-3}$. In contrast to the lasers grown by Evans and Razeghi, these wafers did not have any form of InP-regrowth, and would be expected to have inferior thermal performance. In addition, each wafer was processed into multiple varieties, including the previously described 45 μm broad-area devices and the 4-20 μm ridge lasers. Despite the fact that the published QCL design had produced record performance, the grown structures could not match that performance. None of the broad-area devices or 4 μm ridges lased, and the 20 μm ridges had spotty performance. Furthermore, none of the Razeghi-599 devices lased under any conditions, and the best Razeghi-605 could only do so when the ridges were made sufficiently long. Some of the best Razeghi-606 devices could survive CW operation for a short amount of time when cooled to 77 K, but aside from that, none would survive CW performance at any temperature. Nevertheless, a great variety of MIT-Razeghi devices were available to test and compare, and so they proved to be a useful testbed for many types of measurements. Not only could the transmission experiment be performed, but so could the external Fourier-Hakki-Paoli experiment, internal Fourier-Hakki-Paoli experiment, and photovoltaic spectral measurement.

5.1.1 Absorption measurement

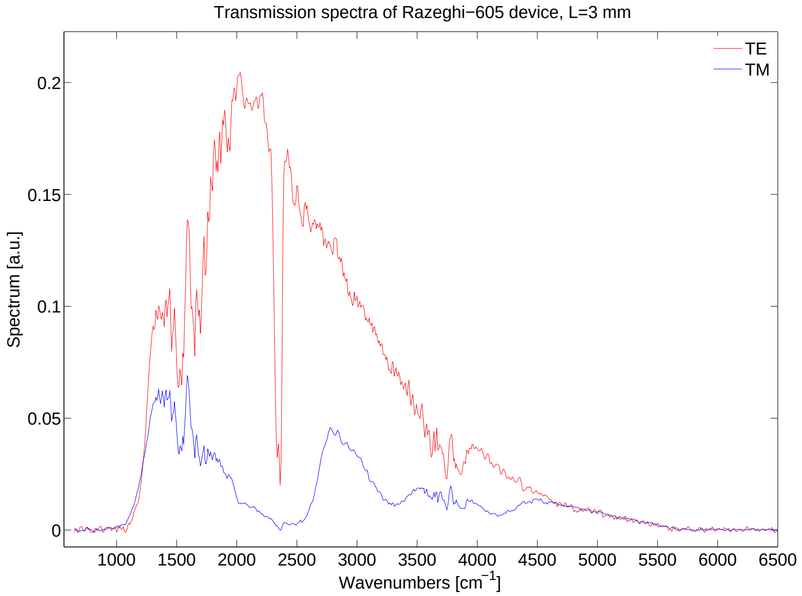


Figure 5-2: Transmission spectra of Razeghi-605 laser (20 μm ridge, 3 mm long)

To start with, consider the results of an absorption experiment. Figure 5-2 contains the polarization dependence of light transmitted through a lasing Razeghi-605 device. (For ease of reading, all TM-polarized spectra are denoted in blue, and all TE-polarized spectra are denoted in red.) Once again, all transmission drops off past 5600 cm^{-1}, a result of absorption of high-energy light by the InGaAs waveguide layers. At the other end of the frequency range, below 10 μm, the waveguide enters cutoff in both polarizations, and no light is transmitted. Incidentally, electromagnetic simulations had predicted that TM-polarized light could support surface plasmons in the 10-20 μm range, but because the device measured was so long, those modes have presumably decayed away.

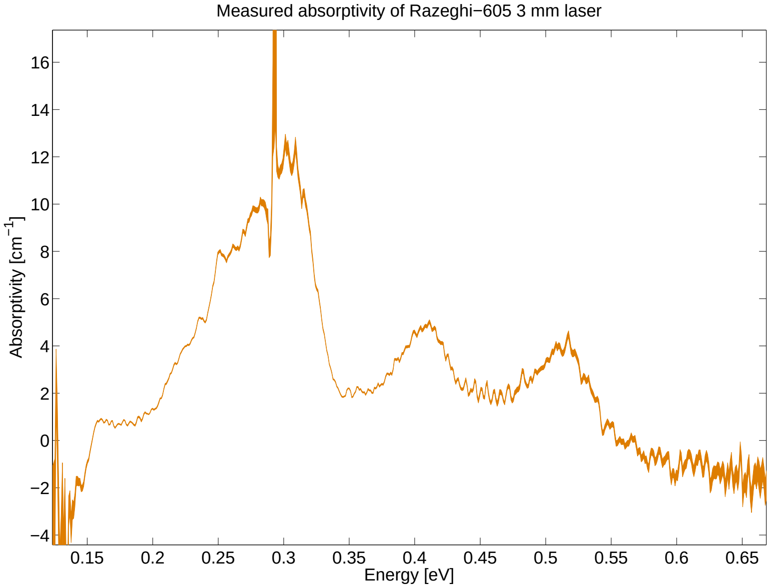


Figure 5-3: Measured absorption spectrum of Razeghi-605 laser (20 μm ridge, 3 mm long)

It is also evident that several absorption features are visible on the TM spectrum. Figure 5-3 plots the absorption calculated from the data contained in 5-2. In order to give the reader an indication of the frequency-dependent signal-to-noise ratio, all absorption spectra will be plotted with a finite thickness, with the vertical extent of the plot representing the calculated spectrum plus and minus one standard deviation. (A cursory examination reveals that the noise indeed increases whenever the signal of either spectra becomes small, as it does near the edges of the plotted range and near the carbon dioxide line at 0.3 eV.)

Unfortunately, the measured data for this device does not agree terribly well with the simulated data, shown in Figure 5-4. First of all, it must be stated up front that in order to get the transition energies to be as close as they appear to be, the well effective

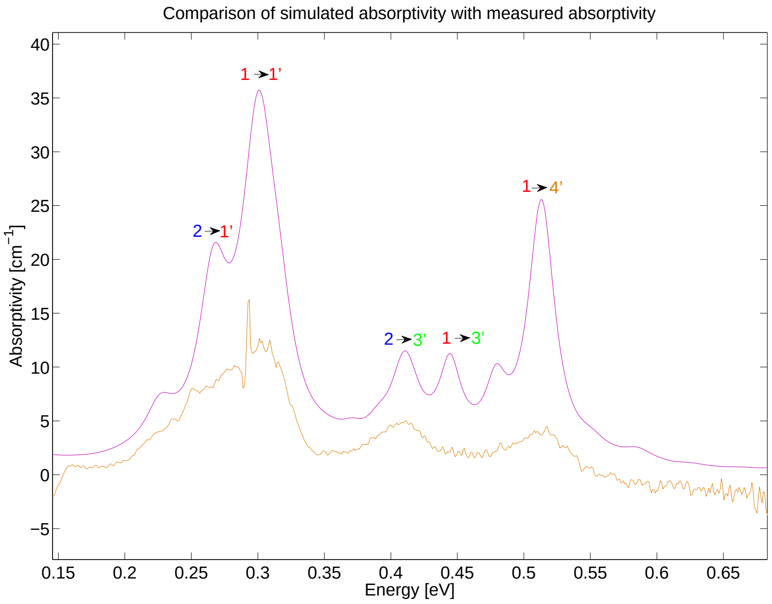


Figure 5-4: Comparison of measurement and simulation of Razeghi-605 laser (20 μm ridge, 3 mm long). Linewidths have all been fixed at 30 meV.

mass had to be adjusted from its tabulated value of 0.039 all the way up to 0.048. However, as mentioned in Chapter 3, this is within the error quoted by Vurgaftman et al., and is acceptable (if not desirable). The more pressing concern is the fact that each measured absorption strength is considerably lower than its predicted value. In terms of total integrated area, the measured data contains only 46.8% of the simulated data. Given the large disparity in area, the most likely explanation for this is that the oxygen traps were formed during the growth process, effectively lowering the doping of the structure and uniformly depressing the peak heights. If this were truly the case, it would also go to explain why this series of devices had spotty performance.

One way to test this hypothesis is to see how the measured absorption strengths change with different doping levels. Figure 5-5 shows the absorptivity measured for a Razeghi-599 device (doped at $1 \cdot 10^{17}\,\mathrm{cm}^{-3}$), a Razeghi-605 device ($2 \cdot 10^{17}\,\mathrm{cm}^{-3}$), and a Razeghi-606 device ($3 \cdot 10^{17}\,\mathrm{cm}^{-3}$). If the dopings were consistent, the absorptivities would be roughly proportional to each other, with the 605 having twice the absorptivity of the 599 and the 606 having three times the absorptivity of the 599. Instead, the opposite trend is visible: as doping increases, transition strength decreases. To make

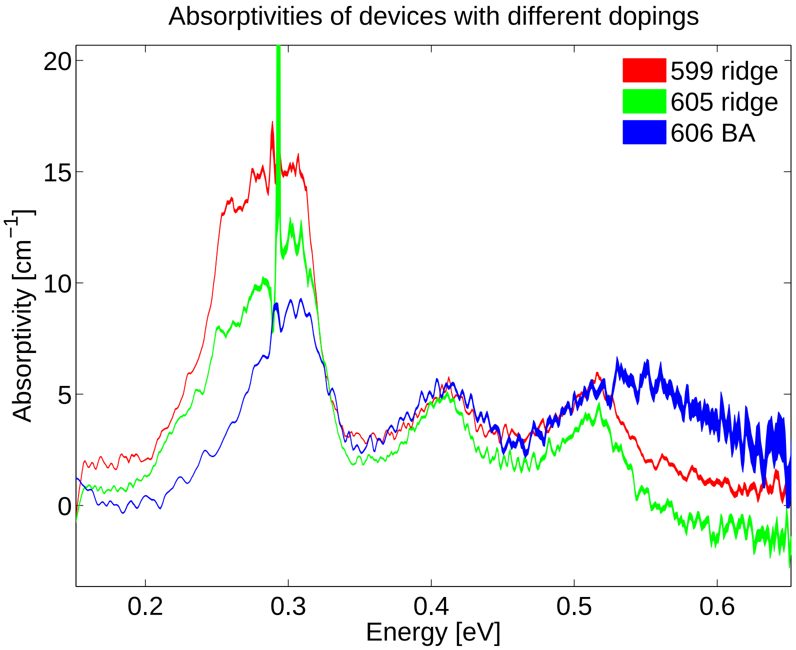


Figure 5-5: Comparison of absorptivities of Razeghi devices with different dopings

matters worse, the *relative* peak height is not even consistent for different dopings, indicating that there existed serious reproducibility problems across different growth runs. On the plus side, none of the measured absorption spectra had total integrated areas that exceeded their predicted values (beyond noise level), which means that whatever the problem was, its only effect was to decrease effective doping.

In fact, the problems intrinsic to the MIT-Razeghi devices go beyond doping. Plotted in Figure 5-6 are the absorptivities of two devices from the *same* wafer and of the same ridge type whose only difference is in length. Though their oscillator strengths are indeed similar, the transition energies differ by about 5%. (Because the devices had already been cleaved into devices before they were characterized, it is impossible to say how far apart they were during the growth process, and the precise spatial variation is unknown.) One explanation for this is that during the crystal growth process, the crystal did not grow with uniform alloy concentration across the surface of the wafer, creating a difference in material parameters. Another possible explanation is that a defect in the crystal forced the strain to somehow be higher some locations than others. However, the x-ray diffraction (XRD) measurement performed

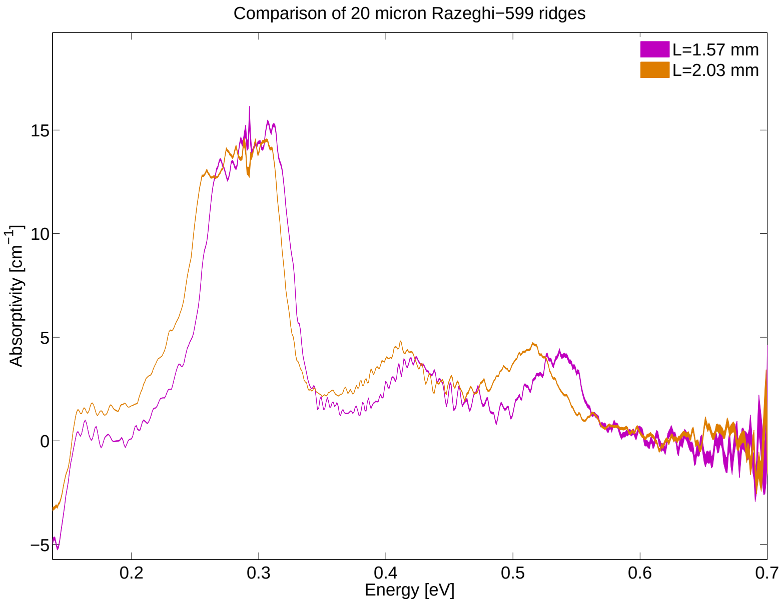


Figure 5-6: Comparison of absorptivities of Razeghi-599 20 μm ridge devices

at Lincoln Laboratory suggested that the QCL module was indeed periodic, and a major defect would manifest itself as an aperiodicity, so this explanation is less likely.

Despite the MIT-Razeghi series' apparent problems, it is still possible to match the experimental data with the simulated data. Even though the measured doping level isn't consistent with its nominal value, it is reasonable to salvage the attempt by scaling the simulated spectrum in such a way that it fixes the doping at a level determined by the total integrated area of the measured spectrum. In the case of the previously-described 3 mm Razeghi-605 ridge, the doping would be changed from $2 \cdot 10^{17}$ cm^{-3} to $9.4 \cdot 10^{16}$ cm^{-3}. Furthermore, lacking *a priori* knowledge of the structure's transition lineshapes, they are assumed to all be Lorentzians, and a best-fit routine is used to determine the linewidths of the major transitions. First, the linewidths of all the transitions are collectively fit to whatever value best minimizes the least-squared error of the measured and simulated data. In the case of the Razeghi-605 ridge, this turns out to be about 40 meV, which is large. Secondly, the linewidths of the most significant peaks are adjusted. The best value of the $1 \rightarrow 1'$ transition is relatively unchanged at 41 meV, and the widths of the $1 \rightarrow 3'$ and $1 \rightarrow 4'$ transitions are increased to 76.3 meV and 71.9 meV, respectively. From a physical point of view, this is palatable: the 1 → 1′ is relatively vertical and occurs between two highly-confined states, while the 1 → 3′ and 1 → 4′ transitions have higher energy, are relatively diagonal, and involve extended states. Therefore, it would be expected that the latter transitions have larger linewidths. When all of this has been done, the calculated spectrum becomes what is shown in Figure 5-7. Clearly, the two spectra now agree fairly well.

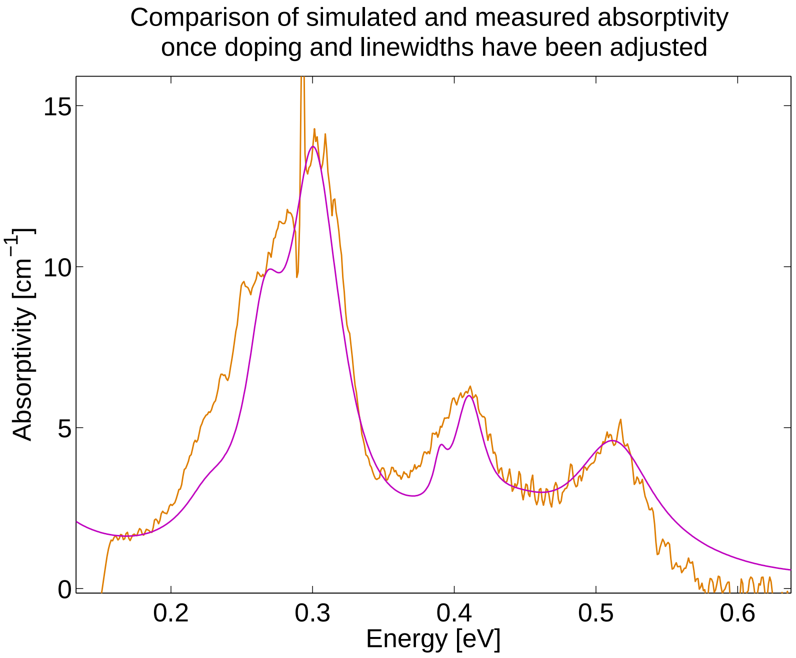


Figure 5-7: Comparison of measurement and adjusted simulation of Razeghi-605 laser (20 μm ridge, 3 mm long)

In order to examine the dependence of coupling effects on ridge width, it is also useful to compare the absorption spectra of a series of devices adjacent to one another and cleaved to the same length (so that on-wafer deviation might be minimized). Figure 5-8 demonstrates the spectra of an 8 μm, a 14 μm, and a 20 μm Razeghi-599 device. (Fortunately, all devices have spectra of comparable shape, and so any variation that occurs on the wafer must occur fairly slowly.) Note that especially at high energies, this series of devices experiences fluctuations that cannot be attributed to detector noise, and the 20 μm device experiences the most severe fluctuations while the 8 μm device experiences the least. Therefore, it is reasonable to attribute them to the presence of higher-order lateral modes (of which the 8 μm device has the fewest and the 20 μm device has the most), confirming the electromagnetic simulation that suggested that a large number of lateral modes should induce a chaotic behavior.

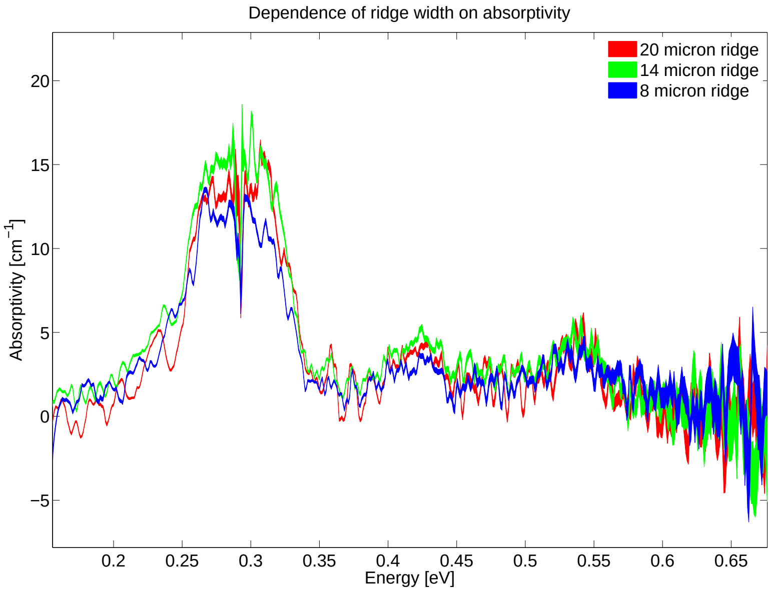


Figure 5-8: Comparison of absorptivities of Razeghi-599 8, 14, and 20 μm ridge devices

5.1.2 Photovoltaic responsivity measurement

Figure 5-9 shows the results obtained from the photovoltaic measurements of two MIT-Razeghi devices, a 20 μm ridge Razeghi-605 device with a length of 3 mm (whose absorption has already been included) and a broad-area Razeghi-606 device with a length of 1.5 mm. (Each has been normalized to the source spectrum to ensure that no frequency has been favored over another.) The Razeghi-605 device shows features similar to its counterpart absorption plotted in Figure 5-3, with the transition energies agreeing very well.

A somewhat surprising development is the difference in apparent lineshapes of the main $1 \rightarrow 1'$ peak on the two devices. Strangely enough, the PV spectrum of the Razeghi-605 ridge (which lases) has less agreement with its measured absorption

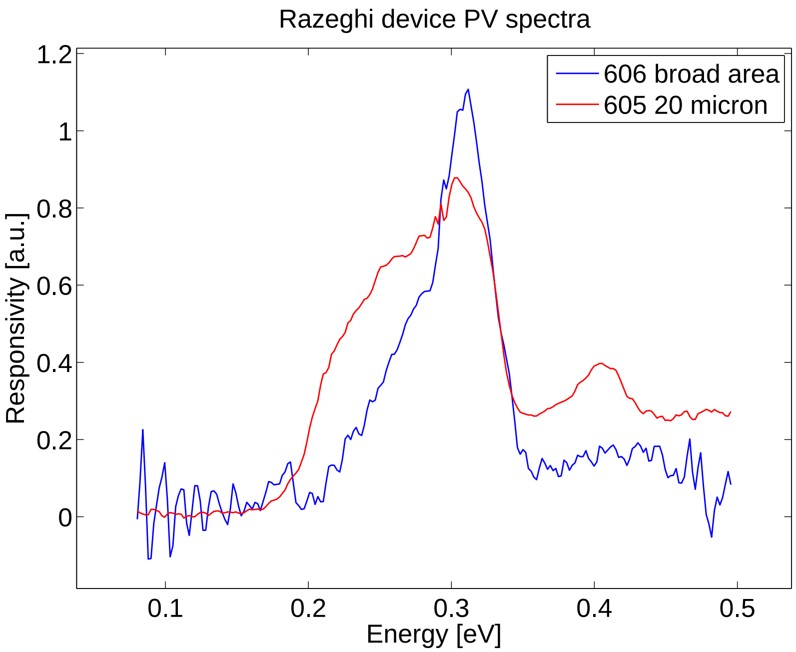
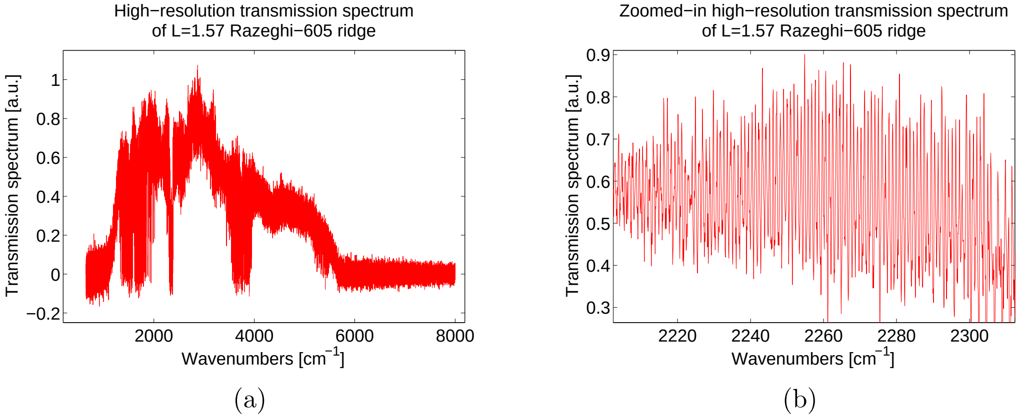


Figure 5-9: Photovoltaic measurements of Razeghi devices

lineshape than does the PV spectrum of the Razeghi-606 broad area device (which does not lase). Apparently, a "saturation" effect occurs with longer devices: as device length is increased, any frequency that has even a slight absorptivity will donate all of its energy to the QCL. However, once this has occurred, no more photovoltage can be induced, and the responsivity saturates at some finite value. More precisely, the responsivity is proportional to $1 - e^{-\alpha(E)L}$ (ignoring the mobility factor of the QCL states' decay). Thus, the long ridge's responsivity has been nonlinearly suppressed near its tallest peaks, thereby appearing to enhance the shorter ones.

5.1.3 Loss measurements

Both internal and external Fourier-Hakki-Paoli loss measurements were performed on MIT-Razeghi devices. Plotted in Figure 5-10 is the high-resolution transmission spectrum of the 1.57 mm Razeghi-599 20 μm ridge whose absorption spectrum is one of those shown in Figure 5-6. Note that a large fraction of what appears to be "noise" on the spectrum actually turns out to be Fabry-Perot fringes upon closer examination. However, because the signal-to-noise ratio is high and the fringe contrast is low,



(a) (b)

Figure 5-10: High-resolution transmission spectrum of 20 μm Razeghi-599 ridge (a) Full spectrum (b) Zoomed-in spectrum

measuring the ratio of the fringes would be an impractical task. When the resulting spectrum is Fourier-transformed, the result becomes Figure 5-11. The position of the second harmonic is exactly twice that of the first, indicating that they are indeed from multiple reflections within the cavity. From their position, one finds that the optical path length of the cavity is 5.19 mm, and assuming that the effective index is 3.3, this means that the cavity is exactly 1.57 mm long. From the harmonic amplitude ratio of 0.1981, one also finds that the room-temperature waveguide losses are 2.64 cm^{-1}. Though this is slightly higher than the simulated value (which was less than 1 cm^{-1}), it does provide good verification that the waveguide is indeed low-loss. The internal Fourier-Hakki-Paoli measurements for this type of ridge structure were already shown in Chapter 4, and by extracting the gain down to the zero-bias situation, one finds that the frequency-dependent losses are about 1.3 cm^{-1}. However, those losses were measured at 77 K, and would be expected to be lower. Therefore, the measurements indicate that the ridge waveguides indeed behave as simulated, and have the properties desired for efficient lasers.

Alas, the same cannot be said for broad-area waveguides, one of whom has a Fourier-transformed transmission spectrum shown in Figure 5-12. Not only does the device have a relatively high loss to begin with (8.06 cm^{-1}, as estimated from the harmonic amplitude ratio), but it also supports a spurious waveguide mode, one whose presence is indicated by the appearance of a peak that does not have an optical path

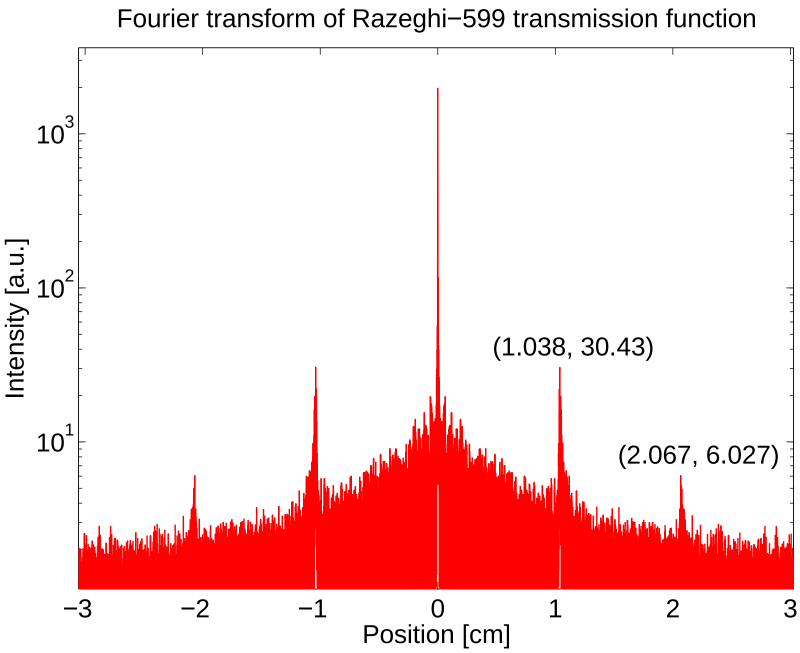


Figure 5-11: Fourier transform of high-resolution transmission spectrum of 20 μm Razeghi-605 ridge

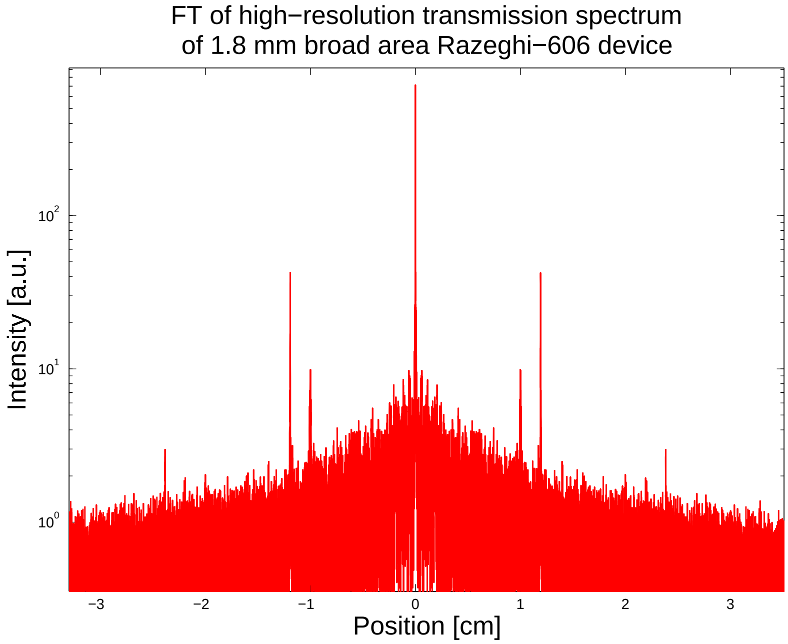


Figure 5-12: Fourier transform of high-resolution transmission spectrum of L=1.8 mm Razeghi-606 broad-area device

length comparable to $n_{\text{eff}}L$. Since the device length is known, the effective index of the spurious peak can be found through simple division, and is found to be about 2.77. This index is considerably lower than the index of any of the III-V materials in the system, and suggests that the mode interacts strongly with the gold contact layer or with the surrounding air. In either case, the mode must also be quite lossy, as its accompanying second harmonic cannot be identified above the noise background. One explanation is that the sloped sidewalls act as a scatterer that mix the transverse and lateral components of light in the waveguide, essentially coupling the main transverse modes of the system to lossy higher-order transverse modes. This process is illustrated with a ray optics diagram, shown in Figure 5-13. In a geometric sense, the sidewalls would gradually add transverse momentum to light in the waveguide, allowing some of it to eventually escape or couple to higher-order transverse modes not predicted by simulation. This would not necessarily prevent guided modes from existing, but would definitely increase their losses.

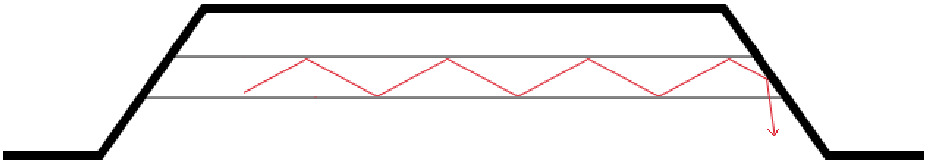


Figure 5-13: Ray optics diagram showing how sidewalls mix momenta

5.2 WSI-Injectorless device

The WSI-Injectorless device was the first in a series of injectorless designs measured, and was provided by the Amman group at the Walter Schottky Institute in order to provide data for the EMIL project. Though only one device was measured, it proved to be a useful verification tool for the transmission experiment, a result of its exceedingly simple band structure (shown in Figure 5-14). The WSI-Injectorless design presented here has proven itself to be very successful, achieving some of the lowest threshold currents ever seen in mid-IR QCLs while lasing at 6.8 μm. [27] This particular device was specifically quite robust, lasing at room temperature even

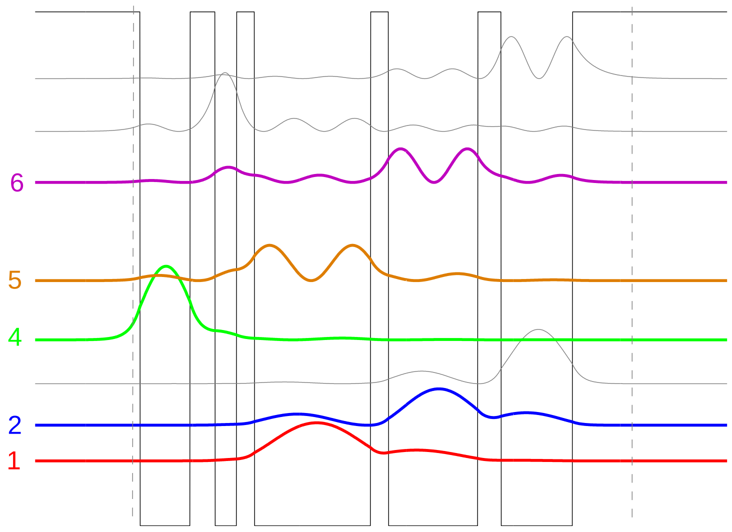


Figure 5-14: Band diagram of WSI-Injectorless structure. Starting from the leftmost dashed line, layer thicknesses are **28**/14/12/10/65/10/50/13/**40**/**28** (in Angstroms, doped layers in bold). Wells are $Ga_{.4}In_{.6}As$, barriers are $Al_{.635}In_{.365}As$.

after suffering substantial mechanical damage. The active region was doped to about $2.6{\cdot}10^{16}\,\text{cm}^{-3}$, and its waveguide structure is very similar to that of the 20 μm Razeghi ridge devices, meaning that the Razeghi electromagnetic simulations are pertinent.

As before, the transmission experiment was performed, and the linewidths were measured by fitting the calculated spectrum to the measured spectrum. After choosing the well effective mass to be 0.0389, the result becomes Figure 5-15. Surprisingly, the degree-of-fit is only fair. The $2 \rightarrow 5$ transition has ended up with a linewidth of just 14 meV, and the $1 \rightarrow 6$ transition has a linewidth of 15 meV. While this in and of itself is not unusual—a well-grown structure could indeed have linewidths that narrow at equilibrium— the linewidth of the $1 \rightarrow 5$ transition is 36 meV, and is considerably higher than the other two. As the former linewidths are due to diagonal transitions and the latter is due to a vertical transition, there is no physical reason that the latter should have a broader lineshape. More importantly, the measured area associated with the $2 \rightarrow 5$ transition is a lot stronger than its simulated counterpart, indicating the state 2 has a higher oscillator strength than would be expected (its

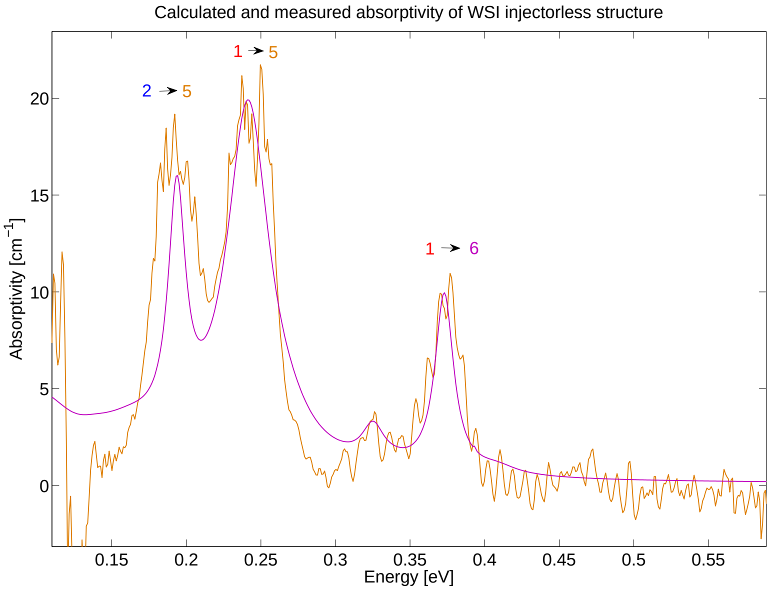


Figure 5-15: Calculated and measured absorptivity of WSI-Injectorless structure. Because the only WSI-Injectorless device that could be measured had neither facet flush with the edge of its mount, the characteristic "ringing" discussed in Chapter 4 is especially evident.

energy and therefore its equilibrium concentration are fixed). This could be a result of the 10 $\AA$ barrier being thinner than expected, in which case state 2 would have stronger overlap with state 5. As an added benefit, this would also have the side effect of making the fitted linewidth appear to be broader.

Loss measurements were also performed on this device, and the results were very interesting. The device itself was cleaved to a length of about 1.8 mm, but during the wire bonding process, a visible crack approximately 1 μm wide formed on the ridge, about 1.1 mm from the rear facet of the device. Though this damage was not catastrophic, the effect of the defect was to create a scattering center midway down the waveguide, one that manifests on the Fourier transform of the high-resolution transmission spectrum, shown in Figure 5-16. Peak B, which is spaced a physical distance of 1.808 mm from the interferogram center burst, is just the usual first harmonic of the waveguide, and is associated with light that has reflected from the rear facet of the device, traveled to the front facet and back, and been emitted. Peak A, which is spaced a physical distance of 1.125 mm from the center burst, is attributed

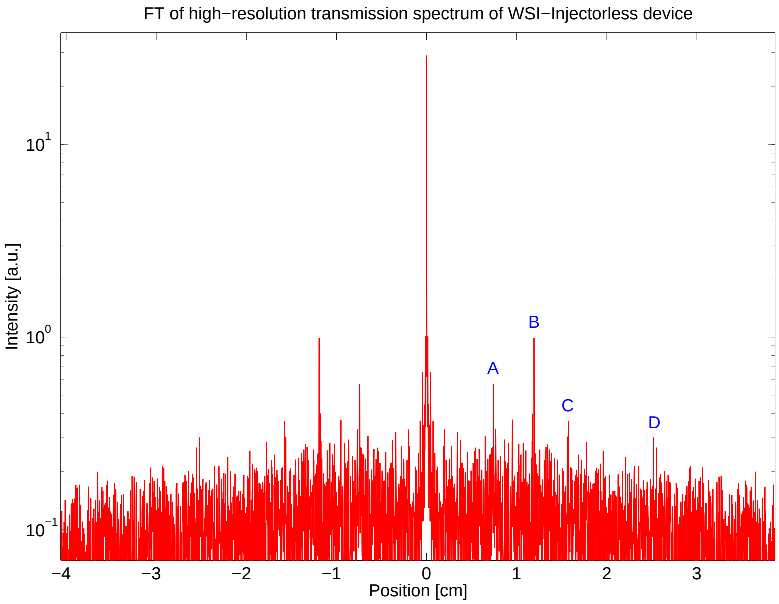


Figure 5-16: Interferogram of damaged WSI-Injectorless device.

to light that reflected off the defect before reaching the front facet, and is therefore spaced closer to the origin than the main harmonic. Peak C, whose distance from the origin is approximately twice that of peak A, is the second harmonic of A; likewise, peak D is the second harmonic associated with peak B. Even with the defect, the ratio of peaks B and D suggests that the total losses are very small: less than 1 cm^{-1}.

Unfortunately, after the WSI-Injectorless device was biased past threshold for a short time, enough heat accumulated in the defect to create localized destruction, preventing further measurements of the device. Still, its clean band structure provided useful information about what a good QCL's transmission properties should be like.

5.3 MIT-Injectorless devices

The MIT-Injectorless devices were also grown as a part of the EMIL program at MIT Lincoln Laboratory, and followed a design philosophy similar to that of the WSI-Injectorless designs. Due to their low voltage defects, they would be expected to have a low threshold and high dynamic range, and because they have fewer subbands than a Razeghi-type device, fewer electrons should idle in non-active transitions. A band diagram showing the structure of the most successful design at zero bias is shown in Figure 5-17. Though the wavefunctions end up being similar to those of the WSI-

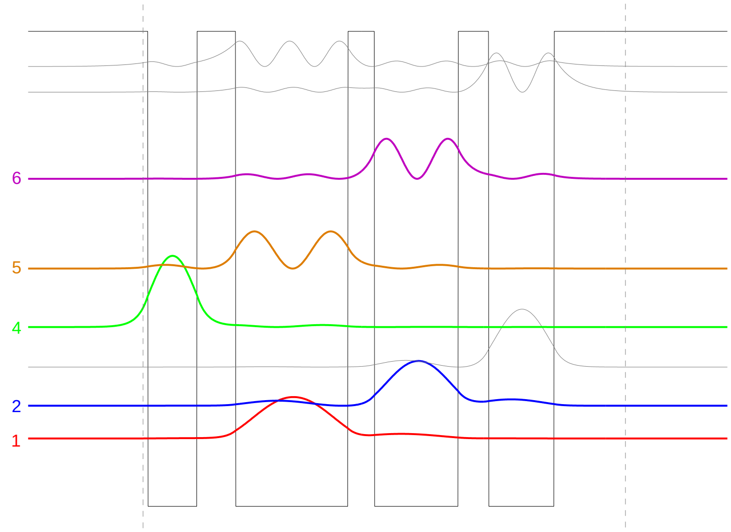


Figure 5-17: Band diagram of MIT-Injectorless structure. Starting from the leftmost dashed line, layer thicknesses are **24**/19/55/13/41/15/32/**32** (in Angstroms, doped layers in bold). Wells are $Ga_{.35}In_{.65}As$, barriers are $Al_{.7}In_{.3}As$.

Injectorless device, there are some key differences. First of all, the alloy compositions of the MIT-Injectorless are more extreme and the structure is more highly strained, in order to obtain a higher emission wavelength of 5.5 μm. Secondly, the design is a four-well design instead of a five-well design, making the structure one of the simplest QCL active regions ever developed. Unfortunately, whatever issues plagued the growth of one of the Razeghi designs remained with the injectorless ones, and only a single device lased, a 20 μm ridge from the Injectorless-708 wafer. Transmission experiments were therefore limited to devices of this type and on this wafer. Figure 5-18 shows the results of the absorption measurement alongside the linewidth-fitted simulation. The well effective mass is found to be 0.0451, and the doping level is determined to be $1.26 \cdot 10^{17}\,\mathrm{cm}^{-3}$, significantly lower than the nominal doping of $4 \cdot 10^{17}\,\mathrm{cm}^{-3}$. Clearly, the trap-like effect that was present in the Razeghi devices is present in these structures as well. The linewidth of the $1 \rightarrow 5$ transition is found

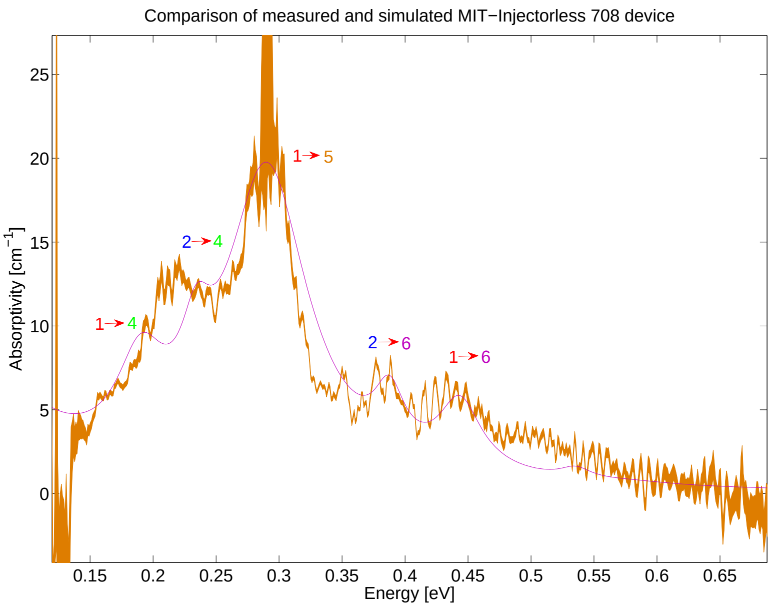


Figure 5-18: Comparison of measurement and simulation of MIT-Injectorless-708 20 μm ridge device

to be about 80 meV, as are the linewidths of the other major transitions. These are comparable to the linewidth obtained through electro-luminescence measurements (about 60 meV), and are even wider than those of the Razeghi structures, indicating that the higher strain of the injectorless designs has worsened the growth quality. Moreover, either the position of the $1 \rightarrow 4$ peak or the $2 \rightarrow 4$ seems to be incorrect, as a peak actually appears in between them on the measurement. Since the $1 \rightarrow 6$ and $2 \rightarrow 6$ transitions appear in the right locations, it is likely that the problem is with the energy of state 4. State 4 is almost completely localized within a single well, and so if that well width was slightly off, it would create a large error in its energy.

As always, it is useful to perform high-resolution transmission measurements to determine the properties of the device's waveguide, and a Fourier-transformed transmission spectrum is shown in Figure 5-19. Once again, spurious modes have appeared on the transmission function. The main harmonics, which correspond to modes completely guided by the active region, appear at positions that correspond to a physical length of 2.05 mm. From these peaks, the waveguide losses are found to be 1.64 cm^{-1}. Likewise, the harmonics belonging to spurious modes have an effective index of about

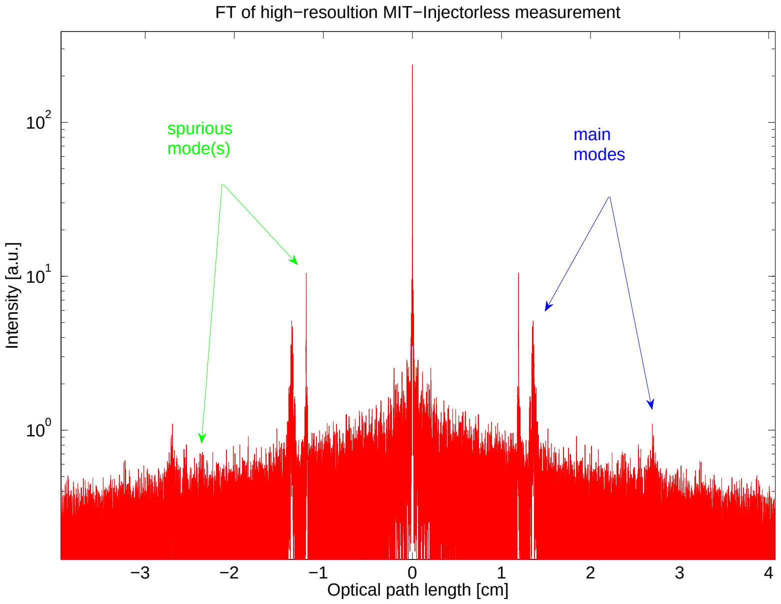


Figure 5-19: High-resolution interferogram of 2 mm MIT-Injectorless device

2.9, and losses of about 7.2 cm^{-1}. Waveguide simulations suggest that no modes should exist at these frequencies and indices, and so it is likely that the model used to estimate the imaginary part of permittivity for lossy materials (i.e., the Drude model) is an insufficient descriptor at these wavelengths. The spurious modes are definitely guided, but cannot be plasmonic in nature due to the fact that they are not extinguished even after traveling a distance of 4 mm.

Chapter 6

Conclusion

In summary, many of the important properties of mid-infrared QCLs have been measured. Through polarization-dependent absorption spectroscopy, a direct measurement of the active region band structure has been performed. This measurement has been used to determine the well effective masses, as well as the real doping concentration and the linewidths of some of the important transitions of different designs. It was also used to prove the existence of growth non-uniformity and trap-like behavior in some of the EMIL structures. Photovoltaic spectroscopy confirmed these results by giving results consistent with them.

High-resolution transmission spectroscopy was used to measure the waveguide losses of several devices, verifying the low-loss predictions of the electromagnetic simulations in some cases but also defying it in others. Through this method, it was shown that different ridge types had very different optical properties, explaining why some could lase and others could not. The measurements also showed the existence of spurious lossy modes that simulation was unable to predict. Moreover, high-resolution spectral measurements provided direct measurements of the gain spectra of devices operating in the amplified spontaneous emission regime, at least in a narrow bandwidth.

While it would have been preferred that biased measurements of devices be performed, technical difficulties with the FTIR used in measurements have prevented this. Further work would include the use of an FTIR with step-scan abilities to measure the performance of pulsed devices, through which properties of active devices such as electron temperature or population inversion could be measured. It would also be useful to obtain a cryostat with a profile thin enough to accommodate high-numerical aperture objectives, in which case parameters at different temperatures could be determined.

Bibliography

[1] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, "Quantum cascade laser," *Science*, vol. 264, pp. 553–556, April 1994.

[2] C. Gmachl, F. Capasso, D. Sivco, and A. Y. Cho, "Recent progress in quantum cascade lasers and applications," *Reports on Progress in Physics*, vol. 64, pp. 1533–1601, 2001.

[3] A. Hsu, "Design of high efficiency mid IR QCL lasers," thesis, Massachusetts Institute of Technology, 2008.

[4] B. W. Hakki and T. L. Paoli, "Gain spectra in GaAs double - heterostructure injection lasers," *Journal of Applied Physics*, vol. 46, pp. 1299–1306, March 1975.

[5] D. Hofstetter and R. L. Thornton, "Theory of loss measurements of Fabry-Perot resonators by Fourier analysis of the transmission spectra," *Optics Letters*, vol. 22, pp. 1831–1833, December 1997.

[6] W. Guo, Q. Lu, Y. Huang, and L. Yu, "Fourier series expansion method for gain measurement from amplified spontaneous emission spectra of Fabry-Perot semiconductor lasers," *IEEE Journal of Quantum Electronics*, vol. 40, pp. 123–129, February 2004.

[7] D. Hofstetter, M. Beck, and J. Faist, "Quantum-cascade-laser structures as photodetectors," *Applied Physics Letters*, vol. 81, pp. 2683–2685, October 2002.

[8] D. G. Revin, L. R. Wilson, J. W. Cockburn, A. B. Krysa, J. S. Roberts, and R. J. Airey, "Intersubband spectroscopy of quantum cascade lasers under operating conditions," *Applied Physics Letters*, vol. 88, pp. 131105–3, March 2006.

[9] E. O. Kane, "Band structure of indium antimonide," *Journal of Physics and Chemistry of Solids*, vol. 1, no. 4, pp. 249–261, 1957.

[10] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures*. Wiley-Interscience, first ed., 1991.

[11] C. Sirtori, F. Capasso, J. Faist, and S. Scandolo, "Nonparabolicity and a sum rule associated with bound-to-bound and bound-to-continuum intersubband transitions in quantum wells," *Physical Review B*, vol. 50, no. 12, p. 8663, 1994.

[12] S. L. Chuang, *Physics of Optoelectronic Devices.* Wiley-Interscience, first ed., August 1995.

[13] B. S. Williams, *Terahertz quantum cascade lasers.* Thesis, Massachusetts Institute of Technology, 2003.

[14] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III–V compound semiconductors and their alloys," *Journal of Applied Physics*, vol. 89, pp. 5815–5875, June 2001.

[15] C. G. V. de Walle, "Band lineups and deformation potentials in the model-solid theory," *Physical Review B*, vol. 39, no. 3, p. 1871, 1989.

[16] M. Lundstrom, *Fundamentals of Carrier Transport.* Cambridge University Press, second ed., November 2000.

[17] H. Callebaut, *Analysis of the electron transport properties in quantum cascade lasers.* Thesis, Massachusetts Institute of Technology, 2006.

[18] H. Callebaut and Q. Hu, "Importance of coherence for electron transport in terahertz quantum cascade lasers," *Journal of Applied Physics*, vol. 98, pp. 104505–11, nov 2005.

[19] A. Wittmann, Y. Bonetti, J. Faist, E. Gini, and M. Giovannini, "Intersubband linewidths in quantum cascade laser designs," *Applied Physics Letters*, vol. 93, pp. 141103–3, October 2008.

[20] S. G. Johnson, P. Bienstman, M. A. Skorobogatiy, M. Ibanescu, E. Lidorikis, and J. D. Joannopoulos, "Adiabatic theorem and continuous coupled-mode theory for efficient taper transitions in photonic crystals," *Physical Review E*, vol. 66, p. 066608, December 2002.

[21] S. Adachi, *Optical Constants of Crystalline and Amorphous Semiconductors: Numerical Data and Graphical Information.* Springer, first ed., August 1999.

[22] H. A. Haus, *Waves and Fields in Optoelectronics.* Prentice Hall, September 1983.

[23] R. J. Bell, *Introductory Fourier Transform Spectroscopy.* Academic Pr, April 1972.

[24] R. Winston, J. C. Minano, P. G. Benitez, N. Shatz, and J. C. Bortz, *Nonimaging Optics.* Academic Press, illustrated ed., 2005.

[25] A. Evans, S. R. Darvish, S. Slivken, J. Nguyen, Y. Bai, and M. Razeghi, "Buried heterostructure quantum cascade lasers with high continuous-wave wall plug efficiency," *Applied Physics Letters*, vol. 91, no. 7, pp. 071101–3, 2007.

[26] A. Evans, J. S. Yu, S. Slivken, and M. Razeghi, "Continuous-wave operation of $\lambda \backsim 4.8\mu$m quantum-cascade lasers at room temperature," *Applied Physics Letters*, vol. 85, no. 12, pp. 2166–2168, 2004.

[27] A. Friedrich, C. Huber, G. Boehm, and M. Amann, “Low-threshold room-temperature operation of injectorless quantum-cascade lasers: influence of doping density,” *Electronics Letters*, vol. 42, no. 21, pp. 1228–1229, 2006.