The Effect of Intracavity Field Variation on the Emission Properties of Quantum Cascade Lasers

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Mansuripur, Tobias S. 2016. The Effect of Intracavity Field Variation on the Emission Properties of Quantum Cascade Lasers. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33840757">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33840757</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
The Effect of Intracavity Field Variation on the Emission Properties of Quantum Cascade Lasers

A thesis presented
by

Tobias Siavash Mansuripur

to
The Department of Physics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Physics

Harvard University
Cambridge, Massachusetts
May 2016
The Effect of Intracavity Field Variation on the Emission Properties of Quantum Cascade Lasers

Abstract

A common and powerful simplification in laser physics is to ignore the spatial dependence of the intracavity field intensity, and instead replace it with its average value. This approach can elucidate many aspects of laser behavior. In this work, however, we examine several problems, both theoretical and experimental, whose understanding requires that the intracavity intensity variation be properly taken into account. We first address theoretically the question of light reflecting from an amplifying slab, a simple problem to pose but one that reveals counterintuitive solutions of the Fresnel equations. These subtleties provide a deeper understanding of negative refraction in nonmagnetic media, amplified total internal reflection, and the perfect lens. Secondly, we fabricate multi-section sampled grating quantum cascade lasers (QCLs) and demonstrate single-mode operation and wide tunability by the Vernier effect. Thirdly, we theoretically investigate how the end mirror reflectivities of a laser affect the output power, and show that power output is reduced when the disparity of the two reflectivities increases. Finally, we demonstrate experimentally for the first time that the transition from single to multi-mode operation in QCLs begins with the appearance of sidebands on the primary lasing mode, separated by tens of free spectral ranges. We explain this state theoretically as the result of the paramet-
ric interaction between the primary lasing mode and the sidebands. The frequency separation of the sidebands and the temporal behavior of the emitted waveform are sensitive to the facet reflectivities. This discovery provides a new pathway toward mid-infrared frequency combs from quantum cascade lasers.
## Contents

Title Page ................................................................. i
Abstract ................................................................. iii
Table of Contents ......................................................... v
Citations to previously published work .............................. vii
Acknowledgments ........................................................... viii
Dedication ................................................................. xiii

1 Introduction ........................................................................ 1
  1.1 Chronological Overview .............................................. 1

2 Fresnel reflection from a cavity with net roundtrip gain ......... 12
  2.1 Abstract ................................................................. 12
  2.2 Introduction ............................................................ 13
  2.3 Theory ................................................................. 13
  2.4 Simulations ............................................................ 17
  2.5 Discussion ............................................................. 20
  2.6 Conclusion ............................................................. 23
  2.7 Supplement: Prescription for R and L Superscripts ................. 24
  2.8 Supplement: Gain Saturation ........................................ 27
  2.9 Supplement: Pulse of light incident on gainy slab with $|\nu| > 1$ .......... 31
  2.10 Supplement: Description of Simulation .......................... 32

3 Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors ........................................... 38
  3.1 Abstract ................................................................. 38
  3.2 Introduction ............................................................ 39
  3.3 Tuning mechanism and simulations ................................. 43
  3.4 Fabrication ............................................................ 48
  3.5 Experiment ............................................................. 50
  3.6 Conclusion ............................................................. 55
# Contents

4 Lasers with distributed loss have a sublinear output power characteristic  
4.1 Abstract ........................................................................................................... 57  
4.2 Introduction ....................................................................................................... 58  
4.3 Distributed Loss Approximation ..................................................................... 63  
4.4 Nonuniform gain saturation  
4.4.1 Ring laser ..................................................................................................... 69  
4.4.2 FP laser ......................................................................................................... 73  
4.5 Other Applications ......................................................................................... 79  
4.6 Conclusion ....................................................................................................... 82  

5 The quantum cascade laser as a self-pumped parametric oscillator  
5.1 Abstract ........................................................................................................... 84  
5.2 Introduction ....................................................................................................... 85  
5.3 Experiment ....................................................................................................... 94  
5.4 Theory ............................................................................................................ 102  
5.4.1 General Framework ..................................................................................... 103  
5.4.2 Population Grating ...................................................................................... 105  
5.4.3 Population Pulsation .................................................................................. 106  
5.4.4 Instability Threshold .................................................................................. 111  
5.5 Discussion ....................................................................................................... 118  
5.6 Conclusion ....................................................................................................... 122  
5.7 Supplement: Calculating intracavity intensity $|\tilde{E}_0|^2$ from measured output power ................................................................. 123  
5.8 Supplement: Gain recovery time vs. Upper state lifetime ......................... 126  
5.9 Supplement: Numerically calculating the instability threshold ............... 127  

6 Theory of parametric gain in lasers  
6.1 Introduction ..................................................................................................... 130  
6.2 Single-mode solution ..................................................................................... 130  
6.3 Frequency mixing due to population pulsations ........................................... 133  
6.3.1 Two-frequency operation ......................................................................... 135  
6.3.2 Three-frequency operation ..................................................................... 138  
6.4 Instability threshold in lasers ....................................................................... 140  
6.4.1 Traveling-wave cavity ............................................................................... 144  
6.4.2 Standing-Wave cavity .............................................................................. 148  

Bibliography ........................................................................................................ 153
Citations to previously published work

Parts of this dissertation cover research reported in the following articles:


Acknowledgments

When I entered graduate school six years ago and was seeking out a lab, the first question I was always asked was “What sort of physics are you interested in?” To this day I am a bit baffled that any 22 year-old can confidently answer this question, as if knowing that AMO, or condensed matter, or particle physics is their pre-ordained path and no other choice will do. Personally, I had no confidence in my ability to choose a problem to work on. I simply wanted to be handed an unresolved problem so that I could start thinking about it, both with theory and experiment. The only thing I was fairly sure of was that the experiments should be of the table-top variety—I like to see the entire experiment in front of me and have control over all the knobs. For providing such an intellectual environment, with the freedom to pursue tangents and unexpected observations, I am indebted to my advisor, Federico Capasso. From what I can tell, there are two key ingredients to his approach to science. First, Federico has the confidence that in an open, collaborative group of researchers, interesting questions will naturally arise. Second, he has an intuition for which of those problems to pursue. Without both of these qualities, we would not have discovered optical parametric oscillation in quantum cascade lasers (QCLs), the main result of this thesis.

When I first joined the group for a rotation in January 2011, Federico assigned me to work with a post-doc, Stefan Menzel, to learn the basics of fabrication and measurement of QCLs. Specifically, Stefan was improving the output power of QCLs with a master-oscillator power-amplifier configuration. Sensing that I was concerned that the project relied more on improved engineering rather than a deeper understanding of physics, Federico assured me that “advances in engineering drive discoveries in
Acknowledgments

physics,” and went on to list numerous historical examples. He told me that with higher power levels, new things could happen which nobody had seen before. I joined the group. Today, I reflect on the prescience of that comment. Our discovery relied on the existence of continuous-wave, high-power QCLs, which have only become available in the last few years thanks to more than a decade of primarily engineering advances.

The path we took to understanding parametric oscillation in QCLs was a circuitous one. I spent about six months interested in the peculiar behavior of one QCL sample, which would randomly jump between a few different emission spectra, even as the current was fixed. When I presented the results at a conference in 2014, Jacob Khurgin made an interesting remark that for the first time led me to consider the importance of the beat note when multiple frequencies are lasing simultaneously. I thought about the idea a little more and told Federico about it when I returned from the conference. He knew immediately that this would be important, and although I was planning to devote more of my time to other experiments, he recruited a Master’s student, Camille Vernet, to help work on this problem. Together, we spent much more time on this problem than I otherwise would have, and after four months it became clear that this work would lead to my thesis. I am beyond grateful for and impressed by Federico’s nudge in the right direction, despite my objections at the time.

I was welcomed into the group by an affable and intelligent crowd, including Romain Blanchard, Patrice Genevet, David Woolf, Francesco Aieta, Mikhail Kats, and Pietro Malara, who made it a pleasure to come to work everyday and quickly made me feel that I’d chosen the right university and the right group. Over the years, I have
Acknowledgments

learned new things by working with many people, including Stefan Menzel, Burc Gok-
den, Laurent Diehl, Christian Pflügl, Patrick Rauter, Yu Yao, Elodie Strupiechonski,
Steve Byrnes, Ilan Gabay, Stefan Kalchmair, Antonio Ambrosio, Guillaume Aoust,
Benedikt Schwarz, Camille Vernet, Bernhard Bohn, and Paul Chevalier. In my first
two years, I learned the most about QCLs from Romain Blanchard, whose easy de-
meanor, willingness to teach, and desire to understand made him a frequent target
of my questions, and I have yet to witness even the tiniest bit of impatience. Thank
you to Alexey Belyanin for all of the scientific discussions in various cities around the
world. I am grateful to Nathalie Picque for hosting me at the Max Planck Institute
for Quantum Optics in Garching, Germany for three weeks, and to Arthur Hipke,
Sam Meek, and Ming Yan for teaching me about frequency combs. Thank you to
Chris Mullaney, Xiomara Forbez, Stacia Zatsiorsky, and Deni Peric for the essential
behind-the-scenes work of keeping the group running.

This work would not have been possible without the generous contributions of
industry collaborators, including David Caffey and Timothy Day at Daylight Solu-
tions, Feng Xie, Chung-en Zah, and Kevin Lascola at Thorlabs, Catherine Caneau
at Corning, and Leo Missaggia, Michael Connors, and Christine A. Wang at MIT
Lincoln Laboratory. Thank you for doing the hard work of developing great QCLs
and maintaining a commitment to do basic science.

The resources provided by Harvard were essential. The Harvard FAS Research
Computing Group runs a superb computing cluster. JD Deng and the entire staff of
the Center for Nanoscale Systems do a fantastic job keeping the machines running,
and I am grateful for all of the training I received.
Acknowledgments

Most importantly, I must recognize my parents. My mom, Anne, has always put my interests ahead of her own. She has made my life easier than I deserve, from childhood to graduate school, even coming to Boston to help me recover from ankle surgery. She has demonstrated by example the importance of having diverse interests, and without her to teach me to swim, take me to soccer, play tennis, and understand the importance of cooking at home, I wouldn’t be who I am. Above all, her compassion has set a standard that I can not live up to, but at least know to strive for. My dad, Masud, has taught and inspired me for as long as I can remember. He taught me to count when I was in a crib. He taught me algebra on a ski lift in Arizona, and binary numbers on a hike in New Mexico. Perhaps most skillfully, he ensured I always viewed math and puzzles as a diversion, not a burden. In college, during long phone conversations about physics homework problems, I started to glean for the first time his approach to physics, and envied his incomparable desire to understand completely. These phone calls continued through graduate school, where, along with technical discussions, he remedied my frustrations and impatience with the reminder that thinking deeply is its own reward, and to not waste energy worrying too much about the point of it all. I do not know what I would have done or where I would be without his guidance, but I know that I would not be as happy. For me, the greatest joy of graduate school has been the opportunity to learn how to think a little bit more like him.

Acknowledgments of Financial Support

I am grateful for the support of the Harvard Physics Department and the Peirce
Acknowledgments

Fellowship, the NSF Graduate Research Fellowship during my first three years, and the support of my advisor, Federico Capasso, through the NSF grants ECCS-1230477 and ECCS-1614631, the DARPA SCOUT program grant W31P4Q-16-1-0002, and support from MIT Lincoln Laboratory.
For my parents
Chapter 1

Introduction

1.1 Chronological Overview

This thesis comprises the major results that I reached during my six years in graduate school, beginning in the Fall semester of 2010. Although in the end, all of the results are somewhat related by the need to understand the intracavity field variation in amplifying media, no effort was made at the outset to ensure that this would be so. In fact, very little effort was made to ensure that there would be any unifying theme at all. For this reason, it may be of interest to to explain chronologically how each of these results developed, if not for the reader, then for myself looking back many years from now.

After my first semester at Harvard, I hadn’t yet found a research group. That winter break, back in Tucson in December, 2010, my dad introduced me to the controversy of amplified total internal reflection (TIR) that he had learned about from an article by Tony Siegman [1]. I told my dad I was frustrated that I didn’t yet
know what I was going to spend my time researching in grad school, so he thought
I might be interested in this. Here’s the crux of the controversy: when a planewave
undergoes TIR, the reflectivity is 1. If the second medium is lossy, then “frustrated”
TIR occurs, and the reflectivity is slightly less than 1 because the second medium
absorbs some of the incident energy. However, if the second medium has gain, then
it is not clear whether the gain should contribute energy to the reflected beam, re-
sulting in a reflectivity greater than 1, or whether the weak transmitted beam should
be amplified ad infinitum, resulting in a reflectivity less than 1 but an exponentially
growing solution in the second medium. It all comes down to a plus/minus-sign am-
biguity when choosing the $k$-vector in the second medium, due to the square root
needed to solve for $k_z$ in the dispersion relation $k_x^2 + k_y^2 + k_z^2 = \omega^2 n^2 / c^2$. My dad
suggested that it would be interesting to consider the behavior of a finite-diameter
beam of light rather than a planewave. Even though a beam is just a superposition of
planewaves, in his experience beams often elucidate aspects of the physics that aren’t
so straightforward to see in the planewave solutions. He also suggested to look at a
finite thickness slab for the second medium rather than a semi-infinite medium, to see
how energy bounced off of the back facet. I started to simulate the beam propagation
in MATLAB, and saw that there was some very counterintuitive behavior and energy
was flowing in unexpected directions. Over time, we realized that this strange behav-
ior was not limited to amplified TIR, but rather occurred whenever the amplifying
slab had net roundtrip gain. Amplified TIR was only a special case of this scenario.
We realized that this solution also explained other phenomena in the literature, such
as so-called negative refraction in amplifying media, as well as Pendry’s perfect lens.
Eventually, we were able to explain how the counterintuitive beam behavior that we were seeing comes about quite naturally from the partial-wave derivation of the Fresnel reflection coefficient. This solution is the inescapable result of applying the Fresnel formalism to Maxwell’s equations; nevertheless, the solution may not uphold causality or be realizable in experiment. However, since the solution had appeared in the literature on many occasions without being properly understood, our goal was just to set the record straight, so that anyone who dealt with the Fresnel behavior of amplifying media was aware of this solution. Our efforts to explain this peculiar solution of Maxwell’s equations are described in Chapter 2.

In January of 2011 I started a rotation in Professor Federico Capasso’s group. I started out by learning the basics: how to electrically pulse a QCL, measure the output power, and measure the emission spectra. Stefan Menzel, a postdoc, had fabricated a variety of QCLs with different kinds of distributed feedback gratings, including three-section sampled grating devices. Fortunately for me, he did not have time to measure the devices himself before moving on to a new job, so I was able to reap the benefits of what he left behind. I did some simulations of the expected emission spectra and measured many thousands of spectra at various driving conditions, and the results are reported in Chapter 3. Later on, Stefan Kalchmair in our group significantly improved these results [2], primarily through the use of antireflection (AR) coatings, and also did more extensive characterization. I lost interest in sampled gratings for two reasons. A few other QCL groups started to look into sampled gratings at the same time, and I don’t enjoy working in an area where others are. Secondly, there didn’t seem to be anything novel to understand, since the
principle had already been demonstrated at telecom wavelengths and we were simply extending the same idea to the mid-IR. Since that time, however, I have understood much better the peculiarities of QCLs, and it is not as straightforward to create a sampled grating QCL as it is a sampled grating bandgap laser, because QCLs are much more difficult to coax into single-mode operation. I also know a bit more now about modeling the intracavity field distribution of each sampled grating mode, and I think there may be some ways to play with the longitudinal overlap factors of the various modes to improve the device properties.

Beginning in 2012, I began to work with Pietro Malara and Romain Blanchard, who had already been working for some time on an external ring-cavity QCL. Spatial hole burning (SHB) was suspected of inhibiting active mode-locking [3] at high pump currents, so Pietro’s idea was to eliminate SHB by placing an AR-AR coated QCL chip into an external ring cavity. Then, by synchronously modulating the current to the device at the roundtrip time of the ring cavity, we hoped that pulsed emission could occur even at high current. The first ring cavity that we built used a QCL chip that could not operate in continuous wave (cw); it could only be electrically pulsed. We characterized the ring cavity’s behavior in [4], and didn’t even attempt active modelocking yet because there was plenty of interesting intrinsic behavior to report, such as the bistable switching between clockwise and counterclockwise modes. Now it was the middle of 2013 and I fully inherited this project. I decided that the cavity’s behavior was so complicated that it should be simplified as much as possible, or else we would never be able to say anything definite about it. In my opinion, the clear first step was to use cw rather than pulsed lasers. Then, the laser would reach a
steady state that we could have some hope of analyzing, rather than dealing with a different emitted waveform after every electrical pulse.

In hindsight, this decision to switch from pulsed to cw lasers strongly influenced the direction of my work, and in a positive way. However, it took a while before I could realize the benefits of working with cw lasers, because I spent many months in the cleanroom developing a cw laser process. I wanted to use a process that did not require buried heterostructure regrowth, in order to avoid the time-consuming step of sending the etched ridges back to the grower for them to do the regrowth. So, I worked on developing an in-house process (that other groups, particularly Professor Razeghi’s, had been using) that relied on thick electroplated gold and flip-chip bonding to improve the thermal dissipation. One big regret I have is that I decided it made sense for me to make my own aluminum nitride submounts, rather than simply pay a company to make them. In the end, it took me about 2 weeks to make the submounts. (I outsourced the AuSn deposition.) And after adding up all the cleanroom costs, I barely saved any money, and in all likelihood my submounts are not as good as commercially produced ones. In any case, by October of 2013 I succeeded in making my own cw laser. It emitted 30 mW at a wavelength of 10 μm at room temperature, while in pulsed mode it emitted around 120 mW of peak power. It was certainly not a record-setting device, but I was proud of it. I placed it in the external ring cavity, took some preliminary spectra, and within a week it died. For reasons I don’t fully understand, I could not get another flip-chipped device to work cw. Subsequent lasers that I processed from other wafers would yield great epi-up devices in pulsed mode, but as soon as I flip-chipped them some leakage channels would open up and
they wouldn’t work well cw. Around this time I started to consider that maybe it was not the smartest idea to try to develop cw lasers by myself, since plenty of people and companies around the world could already do this. I’ll come back to this after discussing some concurrent work I was doing.

When I was studying for my qualifier exam in November of 2013, I reviewed some basic laser derivations, such as the output power as a function of the pumping. The step where the mirror losses are approximated by a distributed loss $\alpha_m$ smoothed throughout the cavity—what I now refer to as the “distributed loss approximation”—bothered me. I didn’t understand why that was an acceptable approximation, done in lieu of calculating the intensity as a function of position in the cavity and properly accounting for the lumped mirror loss at the facets. I set the problem aside so that I could keep studying, but then returned to it after my quals. The differential equations that govern the growth of the intracavity fields are not hard to write down. I first solved the equations numerically and found out that the true output power can deviate significantly from the prediction of the distributed loss approximation. Alexey Belyanin pointed out that in some cases the differential equations could be integrated to yield neat closed-form solutions. Federico pointed out that the analysis was closely related to early laser analyses pioneered by W.W. Rigrod. Whereas Rigrod was concerned with calculating the optimal facet reflectivities to maximize output power, we were pointing out how the LI curve (power vs. current) of a laser became sublinear if the intracavity fields were properly calculated. This theoretical work is presented in Chapter 4. While the sub-linearity of the LI curve is a nice result in its own right, the bigger impact of this work on me was the understanding that the intracavity field
could be strongly tuned by manipulating the end mirror reflectivities. In particular, a Fabry-Perot laser could be turned into more of a traveling-wave laser by placing a high-reflectivity coating on one facet and a low-reflectivity coating (sometimes also referred to as an antireflection (AR) coating) on the other. While Chapter 4 discusses only the role of the intracavity field on the output power, around this time I started to wonder what the effects of various coatings would be on the emission spectrum of a laser. This would turn out to be particularly relevant to our later work on optical parametric oscillation in QCLs, discussed in Chapter 5, although the path we took to get there was not a straight one.

So, in 2014 I was working on the sub-linear LI theory, while trying (and failing) to flip-chip more cw lasers. At the same time, Federico wanted to inject a high-power QCL into a nonlinear chalcogenide fiber to see what kinds of interesting things may come out. We had received three high-power cw QCLs from Mathieu Carras at the III-V Lab for this experiment. While doing a routine characterization of the emission spectra of one of these lasers, I saw something I hadn’t seen before. At a constant current and stabilized temperature, this device would randomly jump between two qualitatively different types of spectra. One spectrum was a familiar QCL spectrum that comprised many adjacent FP modes. The other was a spectrum I hadn’t seen before: most of the energy was concentrated in a single mode, but there were two regions on either side of the central mode, separated by about 7 cm$^{-1}$, where a handful of modes were also lasing. The laser would spend anywhere between a few seconds to a few minutes in one state before jumping to another one. I thought this multistable behavior was interesting, and presented the data at the International
Quantum Cascade Laser School and Workshop in Policoro, Italy, in September 2014. A few days before my talk, I showed some of the data to Professor Jacob Khurgin, and he suggested a possible reason for the existence of the spectrum with a handful of modes separated by a large spectral gap from the dominant single-mode: the QCL tries to avoid any low-frequency amplitude modulation of the intracavity field. Until this moment, I had never thought about how the temporal variation of the field affects the gain seen by each mode. This was the seed that eventually led to our work on optical parametric oscillation in QCLs, which is presented in Chapters 5 and 6.

Originally, I thought of the phenomenon as a temporal analogy of spatial hole burning. In spatial hole burning, two counter-propagating waves of the same frequency create an intensity modulation in space (i.e., a standing-wave) and consequently a population grating, which prevents the maximal extraction of gain from the amplifying medium unless carrier diffusion smooths out the grating. In “temporal hole burning,” two waves at different frequencies form an intensity modulation in time (i.e., a beat note) and consequently a time-varying population inversion, which also prevents the maximal extraction of gain unless a slow population recovery time smooths out the “temporal grating” by preventing the inversion from following the intensity modulation. This was really all I had to go off of, and I thought pursuing this idea would be a tangent that would distract me from working on mode-locking in the external ring cavity. I also wanted to compare the results of the external ring cavity with some intriguing spectra that Romain Blanchard had seen a few years before in QCLs with delayed feedback—essentially an external Fabry-Perot cavity—that showed the persistence of the single-mode state over a large range of current, and also
Chapter 1: Introduction

spectra comprising a handful of modes separated by multiple free spectral ranges.

When I returned from the conference, I told Federico about the idea, and he thought that this would be more interesting than anything else I was working on. I suspect that he thought he could redirect my efforts toward this project by putting a Master’s thesis student, Camille Vernet, to work on the project with me. If that was indeed the plan, it worked out quite well. Rather than spending time on the external ring cavity, Camille and I slowly fumbled our way towards understanding multimode interactions in a two-level system. Starting from the temporal hole burning idea and some numerical simulations of the Maxwell-Bloch equations, we started to piece together how amplitude-modulated and frequency-modulated fields interact with amplifying media. The experimental question that we needed to answer became obvious: as the current is increased past threshold, what is the second mode to start lasing, and what is its frequency separation from the first lasing mode? To do this measurement required cw QCLs with a large operating range of current. The QCLs that I had been able to fabricate weren’t good enough for this. Fortunately for us, in the meantime we had received fantastic cw QCLs from our collaborators at Daylight Solutions, MIT Lincoln Laboratory, and Thorlabs, which we had originally intended to use for external cavity measurements. Instead, these QCLs lent themselves to the much easier measurement of seeing how the spectrum evolved with increasing current. When I performed the measurement, I was ecstatic to see a critical current at which sidebands appeared on both sides of the single-mode, separated by multiple mode spacings.

Despite having the measurement that confirmed our initial suspicion, there was
still plenty that we didn’t understand. We soon found an entire world of literature from the 1960s through the 1980s related to our work: the single-mode laser instability. What we had been referring to as temporal hole burning had been discussed under the name “population pulsations.” Fortunately for us, nobody in the 1980s was discussing QCLs for a very legitimate reason: the QCL was invented in 1994. While earlier work focused on rhodamine dye lasers and the xenon laser, it turns out that the QCL is the ideal system to observe the effects of population pulsations because its picosecond gain recovery time is much shorter than the photon roundtrip time. Also, the previous literature focused heavily on ring cavities, not the Fabry-Perot cavity relevant for QCLs. We also quickly figured out that optical parametric oscillation in microresonators must be closely related to sideband formation in QCLs, although the source of the nonlinearity is different. Therefore, we had a few leads, but there was still a lot to be understood. The big questions that I wanted to answer with the theory were, first, is the 3-mode field amplitude-modulated or frequency-modulated? And second, why do the sidebands appear at pump powers only slightly higher than the threshold? It took about 9 months of work, including chasing a few dead ends, to formulate the theory which appears in Chapters 5 and 6. In the end, I was delighted to have a theory that can answer both of these questions. Best of all, the theory predicted that by putting a high-reflectivity coating on one facet and a low-reflectivity coating on the other, the emitted waveform would be amplitude-modulated and the sideband spacing would be larger than for uncoated lasers. We experimentally confirmed the second prediction, and are currently working on directly confirming the first, although there are already strong hints that it is true. I believe the next step
is to understand what other factors, besides facet coatings, affect the sideband characteristics. The role of group velocity dispersion, temperature, and optical feedback are all worth investigating.

I am excited that we have opened a new window into the world of QCLs, but more excited to have created more questions than we have answered. There are many opportunities for interesting experiments and theories concerning the emission spectra of QCLs, and I hope that a deeper understanding will facilitate the development of frequency comb-based spectroscopy applications.
Chapter 2

Fresnel reflection from a cavity with net roundtrip gain

2.1 Abstract

A planewave incident on an active etalon with net roundtrip gain may be expected to diverge in field amplitude, yet applying the Fresnel formalism to Maxwell’s equations admits a convergent solution. We describe this solution mathematically, and provide additional insight by demonstrating the response of such a cavity to an incident beam of light. Cavities with net roundtrip gain have often been overlooked in the literature, and a clear understanding of their behavior yields insight to negative refraction in nonmagnetic media, a duality between loss and gain, amplified total internal reflection, and the negative-index lens.
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

2.2 Introduction

The Fresnel coefficients govern the reflection and transmission of light for the simplest possible scenarios: at planar interfaces between homogeneous media. Despite this simplicity, some interesting solutions have been discovered only recently, such as 1) the amplification of evanescent waves in a passive, negative-index slab [5–9] and 2) a duality between loss and gain leading to the localization of light in both cases [10–14]. In addition, controversy regarding the proper choice of the wavevector in active media has persisted in relation to the possibility of 3) negative refraction in nonmagnetic media [15–19] as well as 4) single-surface amplified total internal reflection (TIR) [1, 20–24]. It turns out that all four of these cases share a common thread: the presence of a cavity whose roundtrip gain exceeds the loss. In this chapter, we explore in detail the Fresnel solution for such a cavity. We find that its peculiar properties help us to understand the four aforementioned phenomena.

2.3 Theory

To begin, we establish a convention that allows us to more clearly discuss the direction of energy flow. For the single-surface problem, shown in Fig. 2.1(a), the incident wavevector in medium one is $k_1^R = k_x \hat{x} + k_{1z}^R \hat{z}$, and the reflected wavevector is $k_1^L = k_x \hat{x} + k_{1z}^L \hat{z}$, where $k_{1z}^L = -k_{1z}^R$. The superscript R (L) indicates that the wave carries energy to the right (left)—in other words, that the time-averaged $z$-component of the Poynting vector is positive (negative). The real-valued component $k_x$, once determined by the incident wave, is the same for all wavevectors in the system.
the transmitted wavevector, the dispersion relation offers two choices for $k_{2z}$,

$$
k_{2z} = \pm \sqrt{\left(\omega/c\right)^2 \mu_2 \epsilon_2 - k_x^2}, \quad (2.1)
$$

where $\omega$ is the angular frequency, $c$ is the speed of light in vacuum, and $\mu_2$ and $\epsilon_2$ are the relative permeability and permittivity. It is universally agreed that the correct choice for $k_{2z}$ in the single-surface problem is $k_{2z}^{R}$ (i.e., that the transmitted energy flows away from the interface), irrespective of the material parameters or the nature of the incident wave, except possibly in the case of amplified TIR, for which there remains debate. Due to this controversy, let us postulate for now that $k_{2z}^{R}$ is the correct choice in all cases, so that we can unambiguously define the single-surface Fresnel reflection and transmission coefficients

$$
r_{\ell m} = \frac{\tilde{k}_{\ell z}^{R} - \tilde{k}_{m z}^{R}}{\tilde{k}_{\ell z}^{R} + \tilde{k}_{m z}^{R}}, \quad t_{\ell m} = \frac{2\tilde{k}_{\ell z}^{R}}{\tilde{k}_{\ell z}^{R} + \tilde{k}_{m z}^{R}}, \quad \quad (2.2)
$$

where we have generalized the result for incidence medium $\ell$ and transmission medium $m$. For $s$-polarization we have defined $\tilde{k}_{nz} \equiv k_{nz}/\mu_n$, while for $p$-polarization $\tilde{k}_{nz} \equiv k_{nz}/\epsilon_n$. (In cases where both choices for $k_{2z}$ result in no energy flow in the $z$-direction,
such as for evanescent waves in a transparent medium, our prescription is to add a small amount of loss to the slab which will unambiguously distinguish \( k_{R_2} \) and \( k_{L_2} \), then take the limit as the loss goes to zero. See Sec. 2.7 for details.)

We now consider the case of light incident on a cavity, shown in Fig. 2.1(b). The total \( E \)-field resulting from an \( s \)-polarized incident wave in medium one with amplitude \( E_{1R} \) is given by

\[
E_y(x, z) = \begin{cases} 
E_{1R} \exp(ik_x x + ik_{R_1} z) + E_{1L} \exp(ik_x x + ik_{L_1} z) & : z \leq 0 \\
E_{2R} \exp(ik_x x + ik_{R_2} z) + E_{2L} \exp(ik_x x + ik_{L_2} z) & : 0 \leq z \leq d \\
E_{3R} \exp[ik_x x + ik_{R_3}(z - d)] & : z \geq d 
\end{cases} \tag{2.3}
\]

where the time-dependence factor \( \exp(-i\omega t) \) has been omitted. The most direct route to solve for the four unknown wave amplitudes is to enforce Maxwell’s boundary conditions at \( z = 0 \) and \( z = d \), which yields four equations that can be solved for the four unknowns. The resulting reflection coefficient from the slab can be expressed in terms of the single-surface Fresnel coefficients as

\[
r \equiv \frac{E_{1L}}{E_{1R}} = \frac{r_{12} + r_{23} \exp(2ik_{R_2} d)}{1 - \nu} \tag{2.4}
\]

where

\[
\nu = r_{21} r_{23} \exp(2ik_{R_2} d) \tag{2.5}
\]

is referred to as the roundtrip coefficient; the amplitude of a planewave circulating in the slab is multiplied by this factor after each roundtrip in the absence of any sources outside the slab. (Although we explicitly discuss \( s \)-polarized light, our conclusions as well as Eqs. 2.4 and 2.5 hold for both polarization states.) We emphasize that the reflection coefficient given by Eq. 2.4 is a valid solution to Maxwell’s equations for
any value of $\nu$. The roundtrip coefficient $\nu$ has an important physical meaning, and intuitively one would expect three different regimes of behavior when the magnitude of $\nu$ is less than, equal to, or greater than one. The case where $|\nu| < 1$ governs passive slabs (in most but not all cases) and sufficiently weakly amplifying slabs. When $\nu = 1$ the slab behaves as a laser and emits light even in the absence of an incident wave, which manifests itself mathematically as an infinitely large reflection amplitude. The case where $|\nu| > 1$, however, is rarely openly acknowledged [10,19,21].

Probably the reason for the neglect of the $|\nu| > 1$ steady-state solution is the seemingly intuitive assumption that the fields should diverge when there is net roundtrip gain. (See Sec. 2.8 for a discussion of gain saturation and lasing. We also introduce a real example of a cavity with $|\nu| > 1$ whose field stays perfectly finite: the laser subject to self-feedback.) Unfortunately, this incorrect assumption seems to be reinforced by a second well-known solution method for the reflection coefficient that decomposes the reflected wave amplitude $E_1^r$ into a sum over partial waves, yielding the reflection coefficient

$$r = r_{12} + t_{12}t_{21}r_{23}\exp(2ikR d)\sum_{m=0}^{\infty} \nu^m. \quad (2.6)$$

Heuristically, the first term $r_{12}$ (hereafter referred to as the “specular” partial wave) of Eq. 2.6 results from the single-surface reflection of the incident wave at the 1-2 interface, and the geometric series accounts for the contributions to the reflected wave following multiple roundtrips within the slab. When $|\nu| < 1$, the geometric series in Eq. 2.6 converges to $(1 - \nu)^{-1}$, giving the same result as found by matching the boundary conditions in Eq. 2.4. When $|\nu| > 1$, however, the geometric series diverges and the partial wave method suggests that the reflection coefficient is infinite.
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

Intuitively, this divergence seems reasonable, since we expect any light that couples into a slab with $|\nu| > 1$ to be amplified after each roundtrip, and therefore grow without bound. Nevertheless, Eq. 2.4 yields a finite reflection coefficient even when $|\nu| > 1$, so how can we reconcile these two very different solutions?

In fact, the usual heuristic interpretation of the partial wave picture does not tell the whole story, but with a slight modification the partial wave method can be used to find the $|\nu| > 1$ convergent solution. First, one can check that the reflection coefficient given by Eq. 2.4 is invariant under the transformation $k_{2z}^R \equiv k_{2z}^L$, provided $r_{23} \neq 0$. (This can be interpreted simply as a relabeling of the waves $E_2^R \equiv E_2^L$ in Eq. 2.12 that does not affect the final result.) Applying this same transformation to the partial wave sum of Eq. 2.6 [21], we can express the reflection coefficient as

$$r = r'_{12} + t'_{12} t_{21} r_{23} \exp(2i k_{2z}^L d) \sum_{m=0}^{\infty} \nu'^m, \quad (2.7)$$

where the prime indicates the substitution $k_{2z}^R \rightarrow k_{2z}^L$. Because the new roundtrip coefficient, $\nu' = r_{21}' r_{23} \exp(2i k_{2z}^L d)$, is equal to $\nu^{-1}$, in cases where $|\nu| > 1$ the primed partial wave sum of Eq. 2.7 will converge to the reflection coefficient of Eq. 2.4. (This duality between $\nu$ and $\nu^{-1}$ provides a simple mathematical explanation for the loss/gain duality observed by others [10–14]).

2.4 Simulations

The physical implications of the substitution $k_{2z}^R \rightarrow k_{2z}^L$ in the partial wave sum can best be seen by examining the behavior of a “finite-diameter” beam of light incident obliquely on the slab. By numerically superposing a finite number of planewave
solutions to Eq. 2.12 (see Sec. 2.8 for details) with appropriate amplitudes and incidence angles in the range $27.47^\circ < \theta < 32.53^\circ$, we create a Gaussian (to within the sampling accuracy) beam incident on the slab at $30^\circ$ with a full-width at half-maximum (FWHM) beam-diameter of $13.3 \ \mu m$. All media are nonmagnetic, and we choose $\epsilon_1 = \epsilon_3 = 2.25$ and the slab to be an amplifying medium with $\epsilon_2 = 1 - 0.01i$. The free-space wavelength of the beam is $\lambda_0 = 1 \ \mu m$. We can examine the transition from $|\nu| < 1$ to $|\nu| > 1$ simply by varying $d$, since both $|r_{21}|$ and $|r_{23}|$ are less than one (and independent of $d$), whereas $|\exp(2ik_{2z}^{R}d)|$ (and hence $\nu$) increases monotonically with $d$ (because $k_{2z}^{R}$ has a negative imaginary part). A plot of the field $E_y(x, z)$ at one instant of time is shown in Fig. 2.2(a) for $d = 19 \ \mu m$, which was chosen so that $|\nu|$ is slightly less than one for all constituent planewaves of the beam ($0.46 < |\nu| < 0.99$). The arrows overlying the plot point in the direction of the time-averaged Poynting vector within their vicinity, indicating the direction of energy flow in the system, and the incident beam is uniquely identified by the white arrow. The beam behaves as we expect it to: the incident beam strikes the slab near $(x = 0, z = 0)$, giving rise to a specularly reflected beam as well as a refracted beam that ‘zig-zags’ up the slab, which in turn generates a reflected beam in medium one each time it strikes the 2-1 interface. (The field amplitude is plotted on a linear scale, and so the incident beam as well as the specularly reflected beam appear faint relative to the subsequently amplified portions of the beam.) Each of these reflected beams can intuitively be associated with a term of the partial wave expansion of Eq. 2.6–either the specular term or the $m$th term of the geometric series.

In Fig. 2.2(b) all parameters are kept the same except the slab thickness is
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

Figure 2.2: Plots of the field $E_y(x, z)$ at one instant of time for a Gaussian beam (indicated with the white arrow) incident on an amplifying slab for which (a) $|\nu| < 1$ ($d = 19$ μm) and (b) $|\nu| > 1$ ($d = 28$ μm). The black dot indicates the origin of the coordinate system. (For the same material parameters as (b), an incident pulse of light more vividly illustrates the peculiar behavior, as discussed in Sec. 2.9.)

increased to $d = 28$ μm, resulting in $|\nu| > 1$ for all constituent planewaves of the Gaussian beam (1.01 < |\nu| < 2.58). Based solely on the plot of the field amplitude and not on the direction of energy flow indicated by the arrows, it may appear that the incident beam strikes the interface and negatively refracts in the slab, then zig-zags downwards in the $-\hat{x}$ direction, giving rise to many reflected beams in medium one (and transmitted beams in medium three) which emanate from points on the slab
with $x < 0$. Such an explanation was offered for simulations similar to ours [18,19] to attempt to justify negative refraction in an active, nonmagnetic medium. However, by analyzing the Poynting vector we see that the energy in the beam zig-zags up the slab, so this phenomenon is distinct from negative refraction, despite the similarity in the positions of the reflected and transmitted beams. A video of a Gaussian pulse of light (included in the online material, and discussed in detail in Sec. 2.9) with a temporal FWHM of 50 fs and all other parameters identical to those of Fig. 2.2(b) more vividly illustrates that energy flows in the $+x$-direction. We will refer to the field in the slab at $x < 0$ as the “pre-excitation,” so-called because it occurs before the central lobe of the incident beam arrives at the slab. Each reflected beam in Fig. 2.2(b) can be associated either with the specular term $r_{12}'$ or with the $m$th term of the primed partial wave expansion in Eq. 2.7.

2.5 Discussion

The Fresnel solution for a slab with $|\nu| > 1$ is a steady-state harmonic solution, in the sense that if the field distribution presented in Fig. 2.2(b) exists at time $t_o$, then as time is evolved forward the field at each point in space will vary harmonically with frequency $\omega$. The intent of this Letter is not to investigate the causal evolution of the pre-excitation, beginning with the time the excitation source is turned on. We also recognize that the experimental verification of this potential phenomenon will be complicated by factors not included in the Fresnel formalism, such as spontaneous emission, that could lead to instabilities or self-lasing. Experimental work already done on the amplification of evanescent waves [20,25,26] (a regime for which $|\nu| > 1$),
however, has not suffered from either of these problems. We note also that the Fresnel formalism, by implicitly beginning with the time-harmonic subset of Maxwell’s equations, can only elucidate the non-divergent solutions to the full time-dependent equations. There can certainly exist divergent solutions, as demonstrated by finite-element simulations of a wave with a well-defined start-time incident normally on a slab with $|\nu| > 1$ [10].

Rather, we take the pre-excitation behavior demonstrated in Fig. 2.2(b) as the direct, logical, and inescapable consequence of the Fresnel formalism applied to Maxwell’s equations for situations in which $|\nu| > 1$. This solution has surfaced in the literature [1,5–24], sometimes knowingly but often not, but its peculiar properties have not been sufficiently appreciated. Our intent is merely to explore these properties, and explain their relevance to some persistent controversies.

We observe that the specular reflection is given by $r_{12}$ when $|\nu| < 1$ and $r'_{12}$ when $|\nu| > 1$. Since $|r_{12}| < 1$ (in most cases of practical interest) and $r'_{12} = 1/r_{12}$, this means that $|r'_{12}| > 1$, and so the primed partial wave expansion mathematically predicts the amplification of the specularly reflected beam when $|\nu| > 1$. From Fig. 2.2(b) we see that this amplification occurs because the specular beam receives energy from the transmission of the pre-excited field through the 2-1 interface. Another noteworthy feature of this solution is that when $|\nu| \gg 1$ (achieved either by increasing the thickness or gain of the slab, or the incidence angle), the left-propagating wave amplitude $E_{2L}$ becomes much larger than $E_{2R}$. This dominance of the left-propagating wave is a direct (although certainly peculiar) result of the multiple reflections of the pre-excitation at the front and back facets of the slab [7], without which only the
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

right-propagating wave $E_{2R}$ would exist in medium two.

It turns out that TIR from an amplifying slab is well within the regime $|\nu| \gg 1$ (for any reasonable thickness $d$). As $\theta$ surpasses the critical angle for TIR, $|\nu|$ quickly becomes extremely large due to the negatively increasing $\text{Im}(k_{2z}^R)$. (For the parameters used in Fig. 2.2(b), the critical angle is $\theta_c = 41.8^\circ$. For $\theta = 41^\circ$, $|\nu| = 9.34 \times 10^3$, and for $\theta = 42^\circ$, $|\nu| = 1.40 \times 10^{15}$.) Therefore, only the left-propagating wave $E_{2L}$ exists with any appreciable amplitude in the slab, and we emphasize again that this results directly from the multiple reflections of the pre-excitation at both slab facets. Some have argued that even if medium two is semi-infinite, above the critical angle the incident wave excites the wavevector $k_{2z}^L$ in the transmission medium rather than the usual $k_{2z}^R$, resulting in the reflection coefficient $r'_{12}$ and an accompanying amplified specular reflection [20–24]. It seems to us, however, that since the existing experimental results [20,26] can be explained by the slab picture without recourse to the single-surface problem, there is no need for a special TIR-exception to the postulate that $k_{2z}^R$ is the transmitted wavevector in the single-surface problem.

In the case of the negative-index lens [5] (where media one and three are vacuum and medium two has $\epsilon_2=\mu_2=-1$), every incident evanescent wave ($k_x > \omega/c$), for both s and p-polarization, excites a lossless surface plasmon polariton mode [8] on the 1-2 and 2-3 interfaces, resulting in $r_{21} = r_{23} = \infty$ and hence $\nu = \infty$ (despite being a passive medium). Therefore, the convergent solution found by applying the Fresnel formalism to this problem [5,6] is the one described by the primed geometric series in Eq. 2.7 with $\nu' = 1/\nu = 0$. The result is that only the wavevector $k_{2z}^L$ exists in the slab, which describes an evanescent wave that is amplified with increasing $z$, and
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

therefore enables the “perfect lensing” action. In this case, the reflection coefficient from the slab is given by \( r = r'_{12} = 0 \), and our argument indicates that this is the result of multiple reflections within the slab [7], not because the incident evanescent wave is impedance-matched to the slab [6,9]. If loss is introduced to the slab, \( r_{21} \) and \( r_{23} \) become finite but the lens still works well if \( |\nu| \gg 1 \); however, even small losses lead to \( |\nu| < 1 \), causing the decaying wave \( k_{22}^R \) to dominate the amplified wave \( k_{22}^L \), thereby spoiling the perfect lens [8,9].

2.6 Conclusion

We have shown that the convergence of the Fresnel solution for a cavity with net roundtrip gain relies on the existence of the ‘pre-excited’ field, which is a peculiar manifestation of the geometric partial wave series. By elucidating this counterintuitive phenomenon, we hope to have provided a useful alternative perspective for understanding amplified total internal reflection and the negative-index lens. We have also shown a positive-index slab with net roundtrip gain does not negatively refract–however, because the behavior mimics negative refraction insofar as the positions of the reflected and transmitted beams are concerned, the slab could substitute as a negative-index material in certain applications.
2.7 Supplement: Prescription for R and L Superscripts

The energy flux of an \( s \)-polarized planewave whose \( E \)-field is given by

\[
E_y(x,z,t) = E_0 \exp(ik_xx + ik_z z - i\omega t)
\]  

in a medium \((\epsilon, \mu)\) is given by the time-average of the Poynting vector \( \mathbf{S} \equiv \mathbf{E} \times \mathbf{H} \),

\[
\langle \mathbf{S} \rangle = \frac{|E_0|^2}{2\omega \mu_0} e^{-2i\text{Im}(k_z)z} \left( \text{Re} \left[ \frac{k_x}{\mu} \right] \hat{x} + \text{Re} \left[ \frac{k_z}{\mu} \right] \hat{z} \right). 
\]  

Therefore, energy flows in the \(+z\)-direction (‘to the right,’ in our convention) when \( \text{Re}(k_z/\mu) > 0 \), and we denote the wavevector \( k_z \) which satisfies this condition with the superscript R. Any value \( k_z \) for which \( \text{Re}(k_z/\mu) < 0 \) is accordingly labeled with a superscript L. It follows from these definitions that \( k_z^R \) must make an acute angle with \( \mu \) in the complex plane. (For \( p \)-polarized light energy flows in the \(+z\)-direction when \( \text{Re}(k_z/\epsilon) > 0 \), and so \( k_z^R \) makes an acute angle with \( \epsilon \) in the complex plane.)

In cases where \( \langle S_z \rangle = 0 \), we must establish a prescription for resolving the ambiguity in the choice of superscript, which is best illustrated by an example. Consider the case where medium one is a lossless dielectric \((\epsilon_1 > 1, \mu_1 = 1)\), medium two is vacuum, and the incident propagating wave satisfies \( k_x > k_0 \), where \( k_0 \equiv \omega / c \), so that the two choices for \( k_{2z} \) are \( \pm i \sqrt{k_x^2 - k_0^2} \). Both choices for \( k_{2z} \) yield pure evanescent waves and carry no energy along the \( z \)-direction. By adding a small amount of loss to medium two, so that \( \epsilon_2 = 1 + i\epsilon''_2 \) where \( \epsilon''_2 > 0 \), the two choices for \( k_{2z} \) deviate slightly from the imaginary axis as shown in Fig. 2.3(a). Now both waves carry non-zero energy along the \( z \)-direction; the first quadrant solution is \( k_{2z}^R \) (which can be seen
Figure 2.3: Choosing the R or L label for an evanescent wave. (a) The two choices for \( k_{2z} \) are shown for the case of a slightly “lossy vacuum” \( (\epsilon_2 = 1 + i\epsilon''_2 \text{ where } \epsilon''_2 > 0, \mu_2 = 1) \), for the case \( k_x > k_0 \). The first quadrant solution carries energy to the right and is labeled \( k^R_{2z} \), and our prescription is to take the limit \( \epsilon''_2 \rightarrow 0 \) to determine that \( k^R_{2z} \) in the lossless case is along the positive imaginary axis. (b) For a slightly “gainy vacuum” \( (\epsilon''_2 < 0) \) the two solutions for \( k_{2z} \) are in the second and fourth quadrants, and \( k^R_{2z} \) approaches the negative imaginary axis as \( \epsilon''_2 \rightarrow 0 \). The magnitudes of the real and imaginary parts of \( k_{2z} \) in (a) and (b) are approximated using the first order Taylor expansion for small \( \epsilon''_2 \): \( k^2_{0z} |\epsilon''_2| \ll k^2_{x} - k^2_{0} \).

quickly because it makes an acute angle with \( \mu_2 = 1 \) and the third quadrant solution is \( k^L_{2z} \). Our prescription to establish \( k^R_{2z} \) for a true vacuum (i.e., \( \epsilon''_2 = 0 \)) is to take the limit \( \epsilon''_2 \rightarrow 0 \), which yields \( k^R_{2z} \) as the solution along the positive imaginary axis.

Beware that if one adds a small amount of gain rather than loss to medium two, so that \( \epsilon_2 = 1 + i\epsilon''_2 \) where \( \epsilon''_2 < 0 \), then the two solutions for \( k_{2z} \) exist in the second and fourth quadrants as shown in Fig. 2.3(b), and in this case \( k^R_{2z} \) points predominantly along the negative imaginary axis. Thus, we see that the two limiting cases as gain
or loss approaches zero do not yield the same result:

$$\lim_{\epsilon_2'' \rightarrow 0^+} k_{2z}^R = - \lim_{\epsilon_2'' \rightarrow 0^-} k_{2z}^R. \quad (2.10)$$

To have an unambiguous labeling convention for the case $\epsilon_2'' = 0$, we emphasize that one must take the limit as loss approaches zero, which can be different from the limit as gain approaches zero in the case of evanescent waves.

Finally, it is worth noting that this discontinuity in the two limiting cases, apart from being a footnote in establishing a labeling convention, is actually at the heart of the debate over single-surface amplified TIR. When medium two has gain, if one chooses $k_{2z}^R$ as the transmitted wavevector (in accordance with our postulate), then it seems unphysical that as the gain approaches (but does not reach) zero the transmitted wave should still be strongly amplified. To remedy this situation it has been suggested that the correct choice for the transmitted wavevector should be $k_{2z}^L$ when medium two has gain and $k_z > k_0$, so that the transmitted wave decays in the $+z$-direction. We believe, instead, that the discontinuity in the two limits is not as unphysical as it might appear at first: the transmitted wave propagates a large distance in the $x$-direction while barely moving forward in the $z$-direction (since $k_x \gg \text{Re}(k_{2z}^R)$), so the large gain in the $z$-direction is actually a result of the long propagation distance along the $x$-direction. Far more unphysical, in our opinion, is the decision to switch the transmitted wavevector from $k_{2z}^R$ when $k_x < k_0$ to $k_{2z}^L$ when $k_x > k_0$. All of these arguments aside, however, the purpose of our paper has been to demonstrate the Fresnel mechanism by which the specularly reflected beam from a finite-thickness slab is amplified, both below and above the critical angle.
2.8 Supplement: Gain Saturation

Physicists familiar with the principles of lasers should be rightfully wary of a steady-state solution with roundtrip coefficient $|\nu|$ greater than one. When an active medium is pumped strongly enough to generate a sufficiently large population inversion to yield $|\nu|$ greater than one, light initially generated by spontaneous emission in the cavity will be amplified after each roundtrip. However, the field amplitude does not grow without bound— as the field gains strength the upper state lifetime is reduced by stimulated emission, which causes the population inversion to decrease to a level such that $\nu = 1$, resulting in steady-state lasing. This gain reduction with increasing field amplitude is known as gain saturation. In a laser, therefore, the situation $|\nu| > 1$ is only a transient state. It clearly cannot be a steady-state solution, because the field would grow without bound.

The situation changes when we allow an incident wave to strike the active medium, as we do in this paper. Note that $\nu$ is defined as the roundtrip coefficient in the absence of an incident wave; that is, the reflectivity $r_{21}$ is calculated by assuming that there is no wave in medium one arriving at the cavity. To account for the incident wave, we can define an effective facet reflectivity at the two-one interface $r_{21}^{\text{eff}} \equiv E_2^R / E_2^L$. Furthermore, we can define an effective roundtrip coefficient in the slab which replaces $r_{21}$ with $r_{21}^{\text{eff}}$, that is, $\nu^{\text{eff}} = r_{21}^{\text{eff}} r_{23} \exp(2ik_2Rz)$. We emphasize that every possible steady-state solution to the problem under consideration, whether the slab is passive or active, and whether there is an incident wave or not, satisfies the condition $\nu^{\text{eff}} = 1$. This is a fundamental property of steady-state solutions: the field in the slab must regenerate itself after every roundtrip, once all sources and sinks have
Figure 2.4: Schematic of a laser with self-feedback. A mirror of reflectivity $r_{\text{mir}}$ allows light emitted by medium two to be re-injected. The feedback alters the laser threshold, and can be modeled by an effective facet reflectivity $r_{21}^{\text{eff}}$, which differs from the bare facet reflectivity $r_{21}$ in the absence of feedback.

been accounted for. Therefore, in situations where $|\nu| > 1$, the incident wave must, upon transmission into medium two, interfere destructively with the circulating field in the slab so that $|r_{21}^{\text{eff}}| < |r_{21}|$, which ultimately forces $\nu^{\text{eff}}$ toward 1. In summary, when there is no incident wave the situation $|\nu| > 1$ is temporary because the field will grow until gain saturation (a nonlinear effect) forces the $\nu = 1$ solution. With an incident wave, a linear steady-state solution is possible even when $|\nu| > 1$ because of the reduction in the effective facet reflectivity $r_{21}^{\text{eff}}$, which prevents the unbounded growth of the fields so that one does not have to rely on gain saturation to avoid a nonphysical divergence.

An example that illustrates well the distinction between $\nu$ and $\nu^{\text{eff}}$ is a laser subject to self-feedback, shown in Fig. 2.4. In this case, the laser (medium 2) emits light in a direction normal to both of its facets, and a mirror of reflectivity $r_{\text{mir}}$ placed a distance $L$ from the left facet reflects some of the laser output and creates the wavevector $k_{1}^{R}$, which is then reinjected into the laser. (To be clear, this is
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

different from the situation considered in our paper, for which \( k_1^R \) is generated by an external source.) For a suitable choice of \( L \) and \( r_{\text{mir}} \), the re-injected wave will interfere destructively with the circulating field in the gain medium, effectively reducing the reflection coefficient at the 2-1 interface without altering its phase. Specifically, for \( n_1 = 1 \) and \( n_2 = n'_2 + i n''_2 \), the facet reflectivity in the absence of feedback is \( r_{21} = (n_2 - 1)/(n_2 + 1) \). The feedback from the external mirror will reduce the effective reflection coefficient to \( r_{\text{eff}21} = \alpha r_{21} \), where \( \alpha \) is real and satisfies \( 0 \leq \alpha \leq 1 \), provided \( r_m \) and \( L \) are chosen such that

\[
r_m \exp(i4\pi L/\lambda_o) = \frac{(1 - \alpha)r_{21}}{\alpha r_{21}^2 - 1}. \tag{2.11}
\]

So long as \( |r_{21}| < 1 \) (which is generally the case), it is possible to choose \( r_m \) and \( L \) so that \( |r_m| < 1 \), i.e., the external mirror is a simple, passive component. As an application of Eq. 2.11, consider the case \( n_2 = 1.5 - 0.01i \) and \( \lambda_o = 1 \mu m \). The bare facet reflectivity is \( r_{21} = 0.2 \exp(-0.016i) \). If we want the effective facet reflectivity to be \( r_{\text{eff}21} = 0.1 \exp(-0.016i) \) (half the magnitude but the same phase as the bare facet reflectivity), we choose \( \alpha = 0.5 \) and find that we should place a mirror of reflectivity \( r_m = -0.102 \) a distance \( L = 0.499 \mu m \) from the left facet.

Because the external mirror reduces the effective facet reflectivity so that \( |r_{\text{eff}21}| < |r_{21}| \), the effective roundtrip coefficient \( |\nu^{\text{eff}}| \) will be less than \( |\nu| \). When medium two is pumped beyond the lasing threshold, \( \nu^{\text{eff}} \) will be clamped to 1, which means that \( |\nu| \) will be greater than 1. There also exists a subthreshold pumping regime for which \( |\nu| > 1 \). The example of the laser with self-feedback demonstrates that an active cavity subject to an incident field can have a roundtrip coefficient \( \nu \) with magnitude greater than 1, and that this situation is neither transient nor unstable.
In the above example, there is always a well-defined phase relationship between the field circulating in the laser and the re-injected field. This will not necessarily be true when the incident wave is generated by an external source: spontaneous emission, being a stochastic process, can give rise to a field within medium two that has no well-defined phase relationship with the incident wave. Whether this leads to instabilities remains to be decisively answered, but we note that the existing experimental evidence in the \(|\nu| > 1\) regime has not, to our knowledge, detected any instabilities [14, 20, 21]. It is also worth commenting on an important difference between spontaneously-emitted cavity-photons traveling parallel to the surface-normal of the cavity facets versus those emitted at an oblique angle to the optical axis. For photons emitted parallel to the surface-normal, the slab is clearly a cavity that provides feedback, as the light retraces its path on every roundtrip. However, the roundtrip coefficient is a function of incidence angle, and for the material parameters used for Fig. 2.2(b), \(|\nu|\) exceeds one only for incidence angles \(\theta > 27.43^\circ\). In particular, this means that \(|\nu| < 1\) for \(\theta = 0\); therefore, the gain is insufficient for a spontaneously emitted cavity photon in the \(\theta = 0\) direction to cause lasing. In contrast, a cavity photon emitted spontaneously at a large \(\theta\) such that \(|\nu| > 1\) would seem to experience net amplification after each roundtrip. (We say ‘seem to’ because this is the intuitive interpretation to us; however, we mention again that the experiments that have probed the regime of \(|\nu| > 1\) have not detected such problems with spontaneous emission [20, 25, 26].) Should these obliquely-traveling spontaneously-emitted photons prove problematic in a future experiment, however, they will anyway exit the slab at the top and bottom facets, since any real slab must have a finite length.
in the $x$-direction. One could also coat the top and bottom facets with broadband antireflection coatings to facilitate this removal; this way, these spontaneously emitted photons would leave the slab before they are amplified to the point where they saturate the gain. We mention this only as a consideration for a potential experiment that looks for the pre-excitation mechanism. Our intent in this chapter has been to explore the predictions of the Fresnel solution for a slab with $|\nu| > 1$; in the end, only experiments can decide whether this solution is physical.

2.9 Supplement: Pulse of light incident on gainy slab with $|\nu| > 1$

The video file pulse_video.mov included online in the additional material is a time-lapse video of $|E_y|^2$ of a pulse of light, rather than a beam, for the same material parameters as in Fig. 2(b) of the main text: $\epsilon_1 = \epsilon_3 = 2.25$, $\epsilon_2 = 1 - 0.01i$, $\mu_1 = \mu_2 = \mu_3 = 1$, $d = 28 \, \mu$m. The white vertical lines in the video identify the 1-2 and 2-3 interfaces. The incident pulse is $s$-polarized and Gaussian in both space (FWHM = 13.3 $\mu$m) and time (FWHM = 50 fs, or 15 optical cycles). The central wavelength of the pulse is $\lambda_o = 1 \, \mu$m, and the mean incidence angle (i.e., averaged over all constituent planewaves) is $30^\circ$. The size of each video frame is 210 $\mu$m by 150 $\mu$m (height by width). The time elapsed between frames is 10 fs, and the entire video spans 1.22 ps (123 frames total). The field $|E_y|^2$ is plotted on a logarithmic scale covering 3 decades, i.e. red corresponds to the maximum intensity and blue corresponds to intensities less than or equal to $1/1000$th of the maximum.
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

The background in this image is blue, which corresponds to the minimum of $|E_y|^2$, whereas the background in Figs. 2(a) and 2(b) is green because it is the field $E_y$ that was plotted in that case, so that blue corresponded to the maximum negative field.

In the video, one first sees the incident pulse near the bottom left of the screen, traveling up and to the right. The pre-excitation is soon seen in the slab at the bottom of the frame, and the reflected pulse that corresponds to the $m = 1$ term in the primed partial wave expansion leaves the slab and propagates up and to the left in medium one. The pre-excitation in the slab then undergoes one roundtrip as it zig-zags upward, giving rise to a transmitted pulse in medium three followed by the $m = 0$ reflected pulse in medium one. The pre-excitation then makes one more roundtrip, giving rise to another transmitted pulse in medium three, and then approaches the two-one interface at the same time the incident pulse arrives from the opposite side. The two pulses interfere in such a way as to yield an amplified specularly reflected pulse by entirely depleting the energy content of the slab. The fact that the pre-excitation in the slab travels in the $+x$-direction clearly distinguishes this behavior from negative refraction.

2.10 Supplement: Description of Simulation

The $E$-field plots of the Gaussian beams and the video of the pulse were created using MATLAB. The field at each pixel is determined by superposing a large (but of course finite) number of planewave solutions. Therefore, the plots represent analytical solutions to Maxwell’s equations.

As described in the Sec. 2.3, the response of the slab to an incident $s$-polarized
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

A planewave with amplitude \( E_1^R \) and wavevector \( \mathbf{k}^R_1 = k_x \hat{\mathbf{x}} + k_{1z}^R \hat{\mathbf{z}} \) is given by

\[
E_y(x, z) = \begin{cases} 
E_1^R \exp(ik_x x + ik_{1z}^R z) + E_1^L \exp(ik_x x + ik_{1z}^L z) & : z \leq 0 \\
E_2^R \exp(ik_x x + ik_{2z}^R z) + E_2^L \exp(ik_x x + ik_{2z}^L z) & : 0 \leq z \leq d \\
E_3^R \exp[ik_x x + ik_{3z}^R (z - d)] & : z \geq d 
\end{cases}
\] (2.12)

and the time-dependence factor \( \exp(-i\omega t) \) is not explicitly written. The wavevector components \( k_{2z}^R \) and \( k_{3z}^R \) are determined by the dispersion relation

\[
k_{2z}^R = \sqrt{(\omega/c)^2 \mu_\ell \epsilon_\ell - k_x^2},
\] (2.13)

where \( \mu_\ell \) and \( \epsilon_\ell \) are the relative magnetic permeability and electric permittivity constants of material \( \ell \), and the sign of the square root is chosen according to the prescription described in Sec. 2.7. The four unknown wave amplitudes are found by satisfying Maxwell’s boundary conditions to be

\[
E_2^R = \frac{2k_{1z}^R(k_{3z}^R + k_{2z}^R)E_1^R}{(k_{2z}^R + k_{1z}^R)(k_{3z}^R + k_{2z}^R) + \exp(2ik_{2z}^R d)(k_{3z}^R - k_{2z}^R)(k_{2z}^R - k_{1z}^R)}
\] (2.14)

\[
E_2^L = \frac{-2k_{1z}^R(k_{3z}^R - k_{2z}^R)E_1^R}{(k_{2z}^R - k_{1z}^R)(k_{3z}^R - k_{2z}^R) + \exp(-2ik_{2z}^R d)(k_{3z}^R + k_{2z}^R)(k_{2z}^R + k_{1z}^R)}
\] (2.15)

\[
E_1^L = E_2^R + E_2^L - E_1^R
\] (2.16)

\[
E_3^R = E_2^R \exp(i k_{2z}^R d) + E_2^L \exp(-i k_{2z}^R d).
\] (2.17)

To construct the Gaussian beam from the planewave solutions, we begin by expressing \( E_y \) in the \( z = 0 \) plane for a beam traveling parallel to the \( z \)-axis

\[
E_y(x, z = 0) = E_0 \exp \left(-\frac{x^2}{2\sigma_x^2}\right),
\] (2.18)

where \( E_0 \) is the peak amplitude and \( \sigma_x \) is directly proportional to the spatial FWHM

\[
w_x = 2\sqrt{2\ln 2}\sigma_x.
\] (2.19)
By Fourier transforming and subsequently inverting the transform, the field can equivalently be written as an integral in k-space,

\[ E_y(x, z = 0) = \int_{-\infty}^{\infty} dk_x E_1^R(k_x) \exp(i k_x x), \quad (2.20) \]

where

\[ E_1^R(k_x) = \frac{E_0 \sigma_x}{\sqrt{2\pi}} \exp \left( \frac{-k_x^2}{2(1/\sigma_x)^2} \right), \quad (2.21) \]

and the FWHM in k-space is

\[ w_k = 2\sqrt{2\ln 2}/\sigma_x. \quad (2.22) \]

To propagate the beam beyond the \( z = 0 \) plane, we associate with each value of \( k_x \) a component \( k_{1z}^R \) such that the total wavevector obeys the dispersion relation in medium one,

\[ k_{1z}^R(k_x) = \sqrt{(\omega/c)^2 \mu_1 \epsilon_1 - k_x^2}. \quad (2.23) \]

Now the Gaussian beam can be expressed as a function of \( x \) and \( z \) by

\[ E_y(x, z) = \int_{-\infty}^{\infty} dk_x E_1^R(k_x) \exp[i(k_x x + k_{1z}^R z)]. \quad (2.24) \]

At this point, we must approximate the integral in Eq. 2.24 by discretization so that the calculation can be carried out by a computer. We restrict \( k_x \) to a finite sampling width \( w_s \) given by \(-w_s/2 \leq k_x \leq w_s/2\), and sample the beam equidistantly within this region with a total number of samples \( N_s \). The integral in Eq. 2.24 is approximated by the sum

\[ E_y(x, z) = \sum_{k_x = -w_s/2}^{w_s/2} \Delta k_x E_1^R(k_x) \exp[i(k_x x + k_{1z}^R z)], \quad (2.25) \]
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

where

\[ \Delta k_x = \frac{w_s}{N_s - 1}. \] (2.26)

At this point, it is helpful to think of \( E^R_1, k_x, \) and \( k^R_{1z} \) as vectors containing \( N_s \) numerical elements each. To rotate the beam so that it travels at an angle \( \theta \) to the \( z \)-axis, we perform the transformation

\[
\begin{align*}
k_x &\rightarrow \cos(\theta)k_x + \sin(\theta)k_z \quad (2.27) \\
k^R_{1z} &\rightarrow -\sin(\theta)k_x + \cos(\theta)k^R_{1z} \quad (2.28)
\end{align*}
\]

on each element of \( k_x \) and \( k^R_{1z} \). (The Fourier amplitude of each plane-wave \( E^R_1(k_x) \) is unaffected by the rotation in the case of \( s \)-polarized light.) Finally, to displace the waist of the beam to some location \((x_0, z_0)\) in the incidence medium, one must multiply each Fourier amplitude by

\[
E^R_1(k_x) \rightarrow E^R_1(k_x) \exp[-i(k_x x_0 + k^R_{1z} z_0)]. \quad (2.29)
\]

With these redefined values for \( E^R_1, k_x, \) and \( k^R_{1z} \), the sum in Eq. 2.25 is a good approximation to a Gaussian beam traveling at an angle \( \theta \) whose waist is located at \((x_0, y_0)\). The total \( E \)-field at any point in the system is given by

\[
E_{\text{tot}}(x, z) = \begin{cases} 
\text{Real}\{\sum \Delta k_x \left( E^R_1(k_x) \exp[i(k_x x + k^R_{1z} z)] + E^L_1(k_x) \exp[i(k_x x + k^L_{1z} z)] \right)\}, & z \leq 0 \\
\text{Real}\{\sum \Delta k_x \left( E^R_2(k_x) \exp[i(k_x x + k^R_{2z} z)] + E^L_2(k_x) \exp[i(k_x x + k^L_{2z} z)] \right)\}, & 0 \leq z \leq d \\
\text{Real}\{\sum \Delta k_x E^R_3(k_x) \exp[i(k_x x + k^R_{3z} z)]\}, & z \geq d
\end{cases}
\] (2.30)

where \( E^L_1, E^R_2, E^L_2, \) and \( E^R_3 \) are calculated element-wise from \( E^R_1(k_x) \) according to Eqs. 2.14-2.17. The beam plots in Fig. 2.2 are calculated pixel-by-pixel from the
Chapter 2: Fresnel reflection from a cavity with net roundtrip gain

Figure 2.5: A pictorial depiction of the Fourier domain sampling used to generate Fig. 2.2(b). (a) A plot of $|r|$ vs. incidence angle $\theta$ for the case of $\lambda_0 = 1 \mu m$, $\epsilon_1 = \epsilon_3 = 2.25$, $\epsilon_2 = 1 - 0.01i$, and $d = 28 \mu m$. For a beam-width FWHM of 13.3 $\mu m$, the sampling width $w_s = 2w_k$ that we chose restricted the range of incidence angles of the planewaves used to construct the beam to $27.47^\circ < \theta < 32.53^\circ$. In plots (b) and (c), the blue dots overlaying the reflectivity plot indicate the incidence angles of the planewaves that were summed over for (b) $N_s = 501$ samples and (c) $N_s = 2001$ samples. The plot in Fig. 2.2(b) looks visually identical for both choices of sampling values.

The finite nature of the sampling has consequences which must be considered in order to be sure that our results are not affected by numerical artifacts. Firstly, the truncation of the Gaussian beam in $k$-space to the sampling width $w_s$ leads to a convolution with a sinc function in the spatial domain. Therefore, the side-tail of our beam is not truly Gaussian; rather, the envelope of the side-tail is Gaussian but the side-tail itself exhibits periodic sinc-like fluctuations in intensity (which cannot be seen in Fig. 2.2, but can be seen in logarithmic plots which resolve the small
intensities of the side-tail). The sampling width chosen for Fig. 2.2 was \( w_s = 2w_k \) (with \( N_s = 501 \)). We made sure that other choices of the sampling width, \( w_s = 3w_k \) and \( 4w_k \) (with proportionally larger \( N_s \) so that \( \Delta k_x \) remained constant), did not affect the behavior of the plots. Therefore, our conclusions are not affected by the precise value of the sampling width \( w_s \). Secondly, the finite number of samples \( N_s \) implies the spectrum of \( k_x \) values is discrete, so the incident beam is periodic in space. This means that in the plots of Fig. 2.2, there is not just one incident beam but an infinite number of them impinging on the slab, spaced periodically along the \( x \)-axis by a distance \( 2\pi/\Delta k_x = 2830 \) \( \mu \)m. If the sampling is increased from \( N_s = 501 \) to 2001 while keeping \( w_s = 2w_k \) constant (see Fig. 2.5), the distance between adjacent beams increases to 11330 \( \mu \)m, but the plots in both Figs. 2.2(a) and 2.2(b) look identical to the ones with 501 samples. Therefore, 501 samples is sufficient in this case to ensure that the (periodically repeated) beams do not interfere with each other, and that the plot is a good representation of the field of a single beam.
Chapter 3

Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

3.1 Abstract

We demonstrate a three-section, electrically pulsed quantum cascade laser which consists of a Fabry-Perot section placed between two sampled grating distributed Bragg reflectors. The device is current-tuned between ten single modes spanning a range of 0.46 μm (63 cm⁻¹), from 8.32 to 8.78 μm. The peak optical output power exceeds 280 mW for nine of the modes.
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

3.2 Introduction

Most atmospheric trace gases, including CO\textsubscript{2}, NO, CH\textsubscript{4}, and NH\textsubscript{3}, have rotational-vibrational resonances in the mid-infrared (mid-IR) portion of the spectrum, typically between 3 and 16\,\mu m. These resonances provide a unique “fingerprint” for each molecule, which can be identified by an absorption spectroscopy experiment. While a broadband illumination source, such as a Globar, together with a Fourier transform infrared (FTIR) spectrometer can be used for this purpose, the resulting set-up is bulky and contains moving parts. Better portability and sensitivity can be achieved with a narrow linewidth, broadly tunable, high power laser source in the mid-IR, which eliminates the need for a spectrometer. Quantum cascade lasers (QCLs) are uniquely positioned to satisfy these requirements; in fact, their potential for absorption spectroscopy has already been demonstrated and commercialized [27]. Single mode emission is usually achieved by etching gratings into the waveguide to create a distributed feedback QCL (QC-DFB) [28], or by etching the grating into a section directly adjacent to the Fabry-Perot (FP) cavity to form a distributed Bragg reflector (DBR) [29]. Wavelength tuning is achieved by exploiting the temperature dependence of the refractive index [30], either by varying the heat sink temperature or by applying a small dc bias current to cause Joule heating of the active region. In both DFB and DBR devices, the maximum fractional wavenumber tunability $\Delta k/k$ is given by the relative modal refractive index change $\Gamma \Delta n/n$, where $\Gamma$ is the mode overlap factor. Typically, this results in a tuning range of no more than 5 cm$^{-1}$ for current-induced temperature tuning. This spectral coverage is sufficient to detect a molecular absorption line of a gas species, but the large gain bandwidth of QCLs,
which can reach a full width at half maximum (FWHM) of up to 600 cm\(^{-1}\) using specially designed active regions [31–33], allows for more ambitious detection schemes. A wider tuning range allows for multi-line detection, which facilitates calibration. One such scheme is the laser array [34]: a single chip that monolithically integrates thirty two QC-DFBs, each spaced apart spectrally by 9.5 cm\(^{-1}\) for a total spectral coverage of about 220 cm\(^{-1}\). Overlapping the beams in the far field has been successfully demonstrated using wavelength beam combining with a suitable grating [35,36]. The widest spectral coverage to date has been achieved by placing a single multi-stack QCL with a bound-to-continuum active region design within an external cavity in the Littrow configuration [37], and using a rotating grating to provide wavelength-specific feedback to the QCL. The device lases in single mode over a range of 432 cm\(^{-1}\) from 7.6 to 11.4 \(\mu\)m. Truly continuous and mode-hop free tuning in the external cavity configuration requires broadband anti-reflection (AR) coatings with very low reflectivity and precise and active control of both the grating angle and cavity length, which constrain device portability, reliability and ruggedness.

Ideally, the full bandwidth of the QCL gain medium could be accessed continuously, one single mode at a time, in a monolithic device with a single output. This would eliminate the need for beam combining and minimize the amount of driving electronics, thereby providing a cheap, compact, robust platform for trace gas detection. We have made significant progress towards this goal through the use of a three-section QCL that consists of a standard FP gain section placed between two “sampled grating” distributed Bragg reflector (SGR) sections, a strategy that was first pursued at telecom wavelengths [38]. A schematic of the device geometry is shown
in Fig. 3.1(a). Our device is intended as a proof of principle that tuning can be accomplished in an electrically pulsed QCL using sampled gratings; the device differs from traditional telecom sampled grating lasers in that the reflector sections have the same active region as the gain section, there is no phase section, and the facets are not AR-coated. A SGR is essentially a DBR, except that the grating is multiplied by a sampling function. Whereas the DBR is maximally reflective over a narrow band centered at a wavenumber determined by the grating period, the sampling function introduces an additional periodicity which results in a comb of evenly spaced reflectivity peaks centered at the wavenumber determined by the non-sampled grating. A sampled grating incorporated into a single section device operated as a DFB laser can generate multi-wavelength lasing with single mode characteristics at each comb peak; the concept has even been generalized to sampled aperiodic gratings in order to suppress particular comb peaks [39].

True single mode emission and tuning requires the use of two SGRs, and was first described in [40] and demonstrated in [38]; very recently the concept was demonstrated in QCLs using a two-section device operating as two SG-DFB coupled cavities [41]. The two SGRs are chosen to have slightly different comb spacings so that when two of the peaks are aligned, the neighboring peaks are not. In this way, light which travels one roundtrip in the FP cavity sees the largest feedback at the wavelength for which the reflectivity peaks from the two mirrors overlap, and this condition determines the mode with the lowest lasing threshold. Discrete tuning is achieved by heating one SGR section so that its reflectivity spectrum shifts in frequency; for a small shift, two new peaks from each comb will come into alignment as shown in
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

Figure 3.1: (a) Schematic of a typical sampled grating DBR laser, consisting of two sampled grating reflectors (SGR) with slightly different superperiods on each side of a FP cavity. The phase section controls the round-trip phase accumulation, which is critical for continuous tuning, but is not included in our device. Traditionally, the gain medium (orange) is restricted to the gain section and the facets are AR-coated, but in our device the SGRs also have gain and the facets are left uncoated. (b) Reflectivities of the SGRs and basic tuning mechanism. Light traveling in the FP cavity experiences a reflection at both SGRs whose magnitude depends on the frequency of the light. The reflectivities of the two SGRs are depicted as frequency combs (SGR1=red, SGR2=blue) with slightly different spacings due to the difference in the superperiods. When peaks from the two SGRs overlap at a particular frequency, the neighboring peaks do not. Therefore, light at the overlap frequency will experience a large reflection at both SGRs and have a lower lasing threshold compared to other modes. To tune the laser frequency, SGR1 is heated so that its refractive index increases and the reflectivity spectrum correspondingly shifts towards smaller wavenumbers, while SGR2 is not index tuned. The resulting alignment of the two combs is seen to cause a discrete mode hop to a mode with smaller wavenumber. (Not shown: heating SGR2 will instead cause the mode to hop to a larger wavenumber.)

Fig. 3.1(b), causing a discrete jump in the lasing mode. This is known as Vernier tuning by analogy with the Vernier scale, which also utilizes mismatched combs [38]. This strategy overcomes the tuning limitation imposed by $\Delta n/n$ mentioned above.

To achieve quasi-continuous tuning, the two SGR sections must be heated at the same time so that both reflectivity spectra translate together in k-space. To prevent mode hops and achieve truly continuous tuning, a tunable phase section inside the FP cavity is needed to ensure that the wavelength selected by the two SGRs also corresponds to a cavity mode satisfying the roundtrip $2\pi$ phase accumulation condition; this section

42
is not, however, critical for demonstrating discrete tuning. Continuous tuning is still limited by the $\Delta n/n$ factor, and so by setting the comb spacing close to one half of the maximum amount of continuous tuning, it is in principle possible to achieve wide spectral coverage through a combination of discrete jumps and continuous tuning. A more detailed description of the tuning mechanism is presented in the next section.

### 3.3 Tuning mechanism and simulations

Coupled mode theory predicts that light will be strongly reflected from a grating when the wavelength of the light is twice as large as the inverse of a spatial frequency component ($k_{\text{grating}}$) of the grating, which is the Bragg condition. To understand the relation of the Fourier components of the SGR to its design parameters, imagine constructing the SGR by beginning with an infinitely long standard $\lambda/4$ grating, as shown in Fig. 3.2(a). The grating consists of etched and unetched stripes of width $d_e$ and $d_u$ in the laser material and corresponding effective modal refractive indices $n_e$ and $n_u$, where the index contrast is related to the etch depth. Ideally, the widths of the stripes are chosen to be $d_e = \lambda/(4n_e)$ and $d_u = \lambda/(4n_u)$, where $\lambda$ is the vacuum wavelength of interest. The optical period of the grating is then $\Lambda_g = d_e n_e + d_u n_u$. The fundamental spatial frequency of this grating is $1/\Lambda_g$, which is plotted on the axis $k_{\text{grating}}/2$ in Fig. 3.2(b) to emphasize the spectral location of the reflectivity peak at $1/(2\Lambda_g)$ due to the Bragg condition. (Note: we use the standard spectroscopy notation which defines the wavenumber as $k = 1/\lambda$, without the factor of $2\pi$.) To construct the SGR, we first multiply the grating by a top-hat function of width $N_g \Lambda_g$ which reduces the infinite grating to a finite length with $N_g$ periods. The corresponding operation in
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

the spectral domain is a convolution with a sinc function whose width (between first zeros) is $1/(N_g \Lambda_g)$. Second, the finite grating is convolved with a comb function whose spacing $\Lambda_s$ (called the “superperiod”) is to be treated again as an optical path length, which results in a sampled grating of infinite length. The spectrum is correspondingly multiplied by a comb of spacing $1/(2\Lambda_s)$, resulting in a comb of reflectivity peaks modulated by an envelope of width $1/(N_g \Lambda_g)$ centered at $1/(2\Lambda_g)$. The final step, not shown in Fig. 3.2, is to multiply the infinite SGR by a top-hat function of width $N_s \Lambda_s$, so the resulting finite SGR comprises $N_s$ superperiods. The corresponding convolution with another sinc function in the spectrum imparts a linewidth $1/(N_s \Lambda_s)$ to each comb peak. This basic understanding of the SGR parameters and reflectivity spectrum is enough to begin to design a SGR at a particular wavelength and bandwidth, and will prove very helpful in understanding the tradeoffs, such as that between tuning range and side mode suppression ratio (SMSR), that arise in design optimization.

In the remainder of the paper, reflectivity spectra are calculated more accurately by modeling the grating as a one-dimensional dielectric stack with indices $n_e$ and $n_u$ and applying the transfer matrix method.

Finite element simulations (COMSOL) of the waveguide structure were performed to determine the modal refractive indices $n_e = 3.16$ and $n_u = 3.18$ in the etched and un-etched regions of the grating. From this, the grating stripe widths $d_e = 665$ nm and $d_u = 660$ nm were chosen to center the reflectivity spectrum of the SGRs at the peak of the electroluminescence spectrum $k = 1190$ cm$^{-1}$ ($\lambda = 8.4 \mu$m). The number of grating periods $N_g = 11$ in each superperiod was chosen so that the reflectivity envelope did not vary by more than 80% over the expected tuning
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

range. The optical path length $\Lambda_s$ can be expressed in terms of physical distances as $\Lambda_s = N_g(n_e d_e + n_u d_u) + n_u L_{\text{spacer}}$, where $L_{\text{spacer}}$ is the length of unetched region between the grating bursts of two superperiods. The spacings of the two combs, $k_1 = 3.75 \text{ cm}^{-1}$ and $k_2 = 4.40 \text{ cm}^{-1}$, were chosen by setting $L_{\text{spacer1}} = 404.5 \mu\text{m}$ and $L_{\text{spacer2}} = 343.5 \mu\text{m}$. To achieve reasonable reflectivity without the SGRs becoming excessively long, $N_s = 9$ superperiods were included in each SGR. The FP section was 2 mm long.

Figure 3.2: The construction of a SGR and its associated Fourier components can be understood in (a) real space and (b) Fourier space. Beginning with an infinitely long DBR, a SGR is formed by multiplication with a top-hat function and convolution with a comb function. In Fourier space, the fundamental frequency of the DBR is convolved with a sinc function and multiplied by a comb function. The two periodicities $\Lambda_g$ of the grating and $\Lambda_s$ of the sampling manifest themselves as a comb in frequency space of spacing $1/(2\Lambda_s)$ modulated by an envelope centered at wavenumber $1/(2\Lambda_g)$ whose width between the first two zeros is $1/(N_g \Lambda_g)$, where $N_g$ is the number of grating periods.
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

The simulated reflectivity spectra of both SGRs, along with the product of the two reflectivities, are shown in Fig. 3.3 for the grating geometries used in the experiment. From Fig. 3.3(a), we see the mismatch $k_1 - k_2 = 0.65 \text{cm}^{-1}$ was chosen to be slightly larger than the linewidth of the reflectivity peaks, so that when two peaks overlap strongly their nearest neighbors do not. There is a “repeat period” $k_{\text{rep}} \approx k_1 k_2/(k_1 - k_2) = 26 \text{cm}^{-1}$ after which two additional peaks overlap, as seen in Figs. 3.3(b) and 3.3(c), but the reflectivity product is larger for the primary peaks due to the decaying envelope. In the simple Vernier tuning picture, the total amount of tuning cannot exceed $k_{\text{rep}}$ because the peak with the largest reflectivity product will always be within $k_{\text{rep}}/2$ of the center of the envelope. One can increase $k_{\text{rep}}$, and thus the tuning range, by reducing the mismatch $k_1 - k_2$, but at a cost of reducing the SMSR because the nearest neighbors of the overlapping peaks will overlap more strongly. An excellent discussion of these tradeoffs is given in [38].

Figure 3.3: (a) Simulated reflectivity spectra of the two SGRs using the experimental device parameters (with AR coatings assumed on the device facets), showing the two different comb spacing values $k_1 = 3.75 \text{cm}^{-1}$ and $k_2 = 4.40 \text{cm}^{-1}$. (b) Zoomed-out view which shows the repeat spacing at $k_{\text{rep}} = 26 \text{cm}^{-1}$ as well as the envelope modulation of the spectra. (c) Product of the two reflectivities which shows the modes with the largest feedback.

In the simulations shown in Fig. 3.3, AR-coated facets are assumed in order to
more easily identify the relevant comb peaks. In the experimental device, the lack of coatings will lead to a baseline reflectivity of \( |(n_u - 1)/(n_u + 1)|^2 = 0.27 \) which the comb sits upon. The contribution of the grating to the reflectivity (between 0.20 and 0.25 at the comb peaks) is comparable to the facet reflectivity, and therefore crucial in determining the lasing mode. The simulation also assumes that the SGRs are transparent. Because the SGRs contain the same active region as the FP section, this assumption is only strictly valid when each SGR is pumped at the transparency current, when the gain equals the waveguide loss. For smaller (larger) currents, the reflectivity will be smaller (larger) in magnitude. (If the threshold current is exceeded, however, the SGR can act as a DFB laser rather than as a reflector for the FP cavity, which is a separate matter.) In diode SG-DBR lasers, the mirror sections are regrown with passive material, and the index tuning is accomplished by electron plasma and band filling effects which result from current injection [42]. The refractive index of the QCL is much less sensitive to such effects due to the intersubband nature of the transition. Instead, we rely on thermal index tuning resulting from current-induced Joule heating. Including the active region in the SGR sections allows us to take advantage of the QCL material’s ability to generate heat very efficiently.

Discrete mode hopping is achieved by tuning the refractive index of one of the mirror sections, say SGR 1, while the other remains unchanged, which translates the reflectivity spectrum of SGR1 in \( k \)-space and brings the neighboring peaks into alignment. To see why an index change results in a translation, consider a small index change from \( n \) to \( n + \Delta n \) in SGR1 which results from a temperature increase in the InAlAs/InGaAs materials. This causes a small but negligible shift in the peak of
the reflectivity envelope (blue curve in Fig. 3.2(b)). The more significant effect is the change in the optical path length of the superperiod $\Lambda_{s1} \rightarrow \Lambda_{s1} + \Delta \Lambda_{s1}$ which affects the comb spacing $k_1 \rightarrow k_1 - \Delta k_1$ where $\Delta k_1 = \Delta \Lambda_{s1} / (2\Lambda_{s1}^2)$: the entire comb (red peaks in Fig. 3.2(b)) contracts when the SGR is heated with the center of contraction at $k = 0$. The $N$th peak—where $N$ is an integer which counts each peak beginning at $k=0$—shifts by an amount $N \Delta k_1$. In our particular design, we are interested only in the peaks with $N$ between 314 and 320, which corresponds to the small spectral window of 26 cm$^{-1}$ near $k_1 = 1190$ cm$^{-1}$. These few peaks shift by almost the same amount, and so the contraction manifests itself as a translation towards smaller wavenumbers. However, if one were to design a device with a very large tuning range (greater than, say, 200 cm$^{-1}$) then the comb contraction would be relevant and the Vernier tuning mechanism not so straightforwardly applicable. The index dispersion would also need to be accounted for, resulting in unevenly spaced comb lines.

3.4 Fabrication

The QCL wafer used in this experiment was grown by metalorganic chemical vapor deposition (MOCVD) on a conducting InP:S substrate [43]. The lower cladding consists of three layers: an InP layer (InP:Si, $n = 1 \times 10^{17}$ cm$^{-3}$, $d = 3.5$ $\mu$m), a thin grading layer (InGaAsP:Si, $n = 1 \times 10^{17}$ cm$^{-3}$, $d = 30$ nm), and an InGaAs layer (InGaAs:Si, $n = 3 \times 10^{16}$ cm$^{-3}$, $d = 520$ nm). Following the lower cladding, the active region comprises 35 repetitions of the heterostructure layer sequence lattice matched to InP $13/ 47/ 12/ 52/ 11/ 53/ 9/ 17/ 44/ 25/ 36/ 27/ 32/ 27/ 25/ 28/ 21/ 31/ 18/ 34/ 16/ 39/ 15/ 42$, where InGaAs wells are in plain text, AlInAs
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

barriers are in boldface, Si n-doped layers ($n = 4.9 \times 10^{16} \text{ cm}^{-3}$) are underlined, and thicknesses are in units of Angstroms. This is a bound-to-continuum design similar to the one used in [44]. (We have given the layers in the order that they are grown. When the laser is pumped, the electrons travel in the opposite direction, and the 44 Angstrom layer is the injection barrier.) The upper cladding is the same as the lower cladding but grown in reverse order. Finally, a highly doped InP layer (InP:Si, $n = 5 \times 10^{18} \text{ cm}^{-3}$, $d = 500 \text{ nm}$) is grown on top of the upper cladding, followed by 10 nm of very highly doped InP and 20 nm of very highly doped InGaAs. The gain bandwidth was determined by electroluminescence measurements, yielding a peak at $1190 \text{ cm}^{-1}$ and full-width at half-maximum $\approx 250 \text{ cm}^{-1}$.

The wafer processing was initiated by wet etching in an HCl:H$_2$O (1:1) solution to remove the upper cladding layers down to the 520 nm thick InGaAs layer. The sampled gratings were defined by electron beam lithography, and etched to a depth of about 300 nm into the InGaAs layer using reactive ion etching. The original upper cladding was subsequently regrown by MOCVD to create a buried grating. Laser ridges 15 $\mu$m wide were defined by optical lithography and wet etching. An insulating silicon nitride (Si$_3$N$_4$) layer 450 nm thick was deposited using plasma-enhanced chemical vapor deposition. A window was opened into the Si$_3$N$_4$ above the laser ridge by reactive ion etching to allow for electrical contact. An electron beam evaporator was used to deposit 15 nm of titanium followed by 450 nm of gold. Adjacent sections of the device were electrically separated by leaving gaps in the metallization of length 150 $\mu$m, resulting in a contact-to-contact resistance of about 500 $\Omega$ near zero bias. The backside of the sample was then thinned to a thickness of about 150 $\mu$m, and the back
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

Contact was deposited in the same way as the top contact. The lasers were cleaved to a length of 9.1 mm (SGR1=4.1 mm, FP=2 mm, SGR2=3 mm). For simplicity no AR coatings were deposited on the SGR facets, although their use should drastically improve the behavior of the device. Finally, the device was indium mounted epi-side up on a copper heat sink and each section was wire bonded to a gold pad to facilitate testing.

3.5 Experiment

The QCL was mounted on a thermoelectric cooler held at 25 °C. The light output was collected from the facet of SGR2 and sent to a FTIR spectrometer with 0.1 cm⁻¹ resolution. To demonstrate the Vernier discrete tuning experimentally, one of the SGR sections needs to be controllably heated to different temperatures while the FP section and other SGR remain at constant temperature. This was done by controlling the electrical pulse widths and relative delays delivered to the different QCL sections, as shown in Fig. 3.4. A long “tuning pulse” of duration \( t_{\text{tune}} = 3 \mu s \) was delivered to one of the mirrors (SGR1 in the figure). During this time the temperature of SGR1 will rise, but the device will not lase until a short pulse of duration \( t_{\text{lase}} = 50 \text{ ns} \) is applied to the FP and SGR2 sections after a time \( t_{\text{del}} \) relative to the start of the tuning pulse. The delay \( t_{\text{del}} \) was varied from 0 to 2900 ns in increments of 100 ns, with longer delays corresponding to a higher temperature in SGR1 during lasing. It was determined experimentally that \( t_{\text{lase}} = 50 \text{ ns} \) was short enough to prevent the temperature of SGR1 from changing substantially during lasing and therefore prevent intra-pulse mode hops, although 50 ns is long enough to cause some linewidth broadening by
intra-pulse heating. It was necessary to pulse SGR2 concurrently with the FP section in order to reduce absorption losses in SGR2 and thereby lower the lasing threshold to an experimentally accessible value. This was achieved by electrically connecting SGR2 and the FP section with wire bonds and driving both of them with one pulser. While this configuration suffers the drawback that the current through the FP and SGR2 sections cannot be individually controlled, this turns out to not be a problem for discrete tuning. Continuous tuning can in principle be achieved by independent control of the length and delay of the pulse delivered to SGR2, allowing one to heat SGR2 controllably before lasing is initiated by pulsing the FP section. There remains the question of which current magnitudes \( I_1 \) for SGR1, \( I_{FP,2} \) for FP+SGR2 to use to optimize power output and SMSR; since there is no obvious answer a priori, we performed a large parameter sweep over six values of \( I_1 \) (from 0.41 to 1.27 A), seven values of \( I_{FP,2} \) (from 0.93 to 1.96 A), and thirty values of \( t_{del} \) and recorded a spectrum at each of the 1260 driving conditions. In another experiment, the tuning pulse was applied to SGR2, while SGR1 and the FP section were electrically connected. A larger parameter sweep over 9 values of \( I_2 \) (from 0.07 to 1.42 A), 12 values of \( I_{FP,1} \) (from 0.60 to 2.56 A), and 30 values of \( t_{del} \) was performed for a total of 3240 spectra. The spectra were analyzed by an automated software routine to assist in determining the optimal conditions for single mode operation.

Many of the recorded spectra exhibit multimode behavior—typically between two and eight weak modes—which most likely correspond to driving conditions at which none of the reflectivity peaks of the two mirrors strongly overlap within the gain bandwidth. For the driving conditions at which the reflectivity peaks do overlap
Figure 3.4: Experimental approach to achieving discrete tuning. A long tuning pulse of width $t_{\text{tune}} = 3\mu s$ is applied to one of the SGR sections (SGR1 in the figure), during which time the section heats up. A short pulse of width $t_{\text{lase}} = 50\text{ ns}$ applied to the FP section and SGR2 stimulates lasing, and the delay between the start of the two pulses is controlled by $t_{\text{del}}$ to achieve different temperatures of SGR1 during lasing and therefore enable tuning.

well, one mode substantially dominates. In Fig. 3.5(a), we plot the spectra taken at $I_1 = 1.27\text{ A}$ and $I_{\text{FP,2}} = 1.62\text{ A}$ for $t_{\text{del}} = 0, 500, 800, 1700, 2200,$ and $2700\text{ ns}$ for the experiment in which SGR1 is tuned. As SGR1 is heated, the modes hop to the left in steps of $k_2 = 4.40\text{ cm}^{-1}$ as the theory predicts. Small variations around $4.40\text{ cm}^{-1}$ are expected because the mode does not lase exactly at the reflectivity peak resulting from the overlap of the combs of SGR1 and SGR2, but rather at the cavity mode closest to the reflectivity peak which satisfies the roundtrip $2\pi$ phase accumulation condition. For the largest value of delay at $2700\text{ ns}$, we see that the mode has jumped a large distance to the right but ends up $4.40\text{ cm}^{-1}$ away from the mode at $0\text{ ns}$ delay, which is expected due to the restriction of the tuning range to one repeat period. This particular behavior is consistent with simulations of the reflectivity product as
the index change $\Delta n/n$ of SGR1 is varied from 0.25% to 0.52%, as seen in Fig. 3.5(b). The initial detuning of 0.25% was chosen to replicate the experimental data of four mode hops to the left followed by one to the right, and represents a small offset of the actual refractive index compared to the simulated value. Each mode hop requires an index change of about 0.06%, which corresponds to a temperature increase close to 10 K. (This value is estimated using $dn/dT = 2 \times 10^{-4}$, which is calculated based on the temperature tuning of a DFB laser at the same wavelength and made of the same material system, see Fig. 6 in [30]). As can be seen by comparing Figs. 3.5(a) and 3.5(b), the experimentally measured modes appear to be red-shifted about 30 cm\(^{-1}\) relative to the expected peak reflectivity product from the simulations. This is not due to uncertainty in the refractive index, because the clear observation of mode hops at the designed value of 4.40 cm\(^{-1}\) suggests that the index $n_u$, which is largely responsible for determining the mode spacing, is quite close to the simulated value 3.18. Further investigation is required to determine the cause of the red-shift.

It is clear from Fig. 3.5(a) that not all modes show the same SMSR, but it is conceivable that a better SMSR can be achieved at different driving conditions. To test this, the full set of 1260 spectra was searched for single modes with SMSR exceeding 10 dB, which yielded nine distinct single modes (four of which exceed 19 dB SMSR) as shown in Fig. 3.6(a). All but two of the modes emit more than 300 mW of peak optical power. The black arrows point to the six modes at the locations previously seen in Fig. 3.5, spaced apart by 4.40 cm\(^{-1}\) and covering the full repeat period. Interestingly, three additional modes are found between 1170 and 1190 cm\(^{-1}\) and so the total tuning range of 47 cm\(^{-1}\) exceeds the repeat spacing; these additional
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

Figure 3.5: (a) Spectra recorded at various values of $t_{del}$, indicated in the figure, for $I_1 = 1.27$ A and $I_{FP,2} = 1.62$ A, demonstrating the expected mode hops to the left in steps of $k_2 = 4.40$ cm$^{-1}$ as SGR1 is heated. The asymmetric line shape seen in some of the side modes is an artifact from the FTIR spectrometer. (b) Simulation of the normalized reflectivity product of the two mirrors at various values of the fractional index tuning $\Delta n/n$, indicated in the figure, of SGR1. The black arrow for a particular $\Delta n/n$ indicates the wavenumber which sees the highest reflectivity, and gives an approximate indication of the mode most likely to lase if one ignores the small variation in the gain spectrum.

Figure 3.6: (a) Spectra recorded for the largest value of $I_{FP,2}$ tested (1.96 A). It is clear that a complete model of these devices must take into account more than just the reflectivity of the SGR sections. If properly understood, this could open the door to device designs which exceed the repeat spacing in a controllable fashion.

In the experiment for which SGR2 was tuned rather than SGR1, the spectra collected in the parameter sweep yielded ten distinct single modes as shown in Fig. 3.6(b), each exceeding 13 dB in SMSR, covering a spectral range of 63 cm$^{-1}$. The peak power displays a clear envelope modulation, with the modes in the center exceeding 500 mW, and only the mode at the largest wavenumber emitting less than 280 mW. One pair of adjacent modes (at 1154.2 and 1157.6 cm$^{-1}$) exhibits a separa-
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

tion of 3.4 cm\(^{-1}\), while the mean separation between the remaining pairs of adjacent modes is 7.45 ± 0.05 cm\(^{-1}\). This separation is close to twice the expected difference of \(k_1 = 3.75\) cm\(^{-1}\). At present we cannot explain why it appears that half of the expected modes are missing, or why this nearly perfect comb of modes spaced apart by 7.45 cm\(^{-1}\) spans a bandwidth more than twice the repeat period. This large tuning range is more than a factor of 12 greater than that achieved by electrical current tuning in long-wavelength QC-DFB lasers.

Figure 3.6: Experimentally measured single modes, taken at different values of the applied currents and delay, not indicated, and individually normalized. (a) SGR1 is temperature tuned as discussed in the text. The black arrows denote the six peaks at the same locations as seen in Fig.3.5(a), which span the full predicted tuning range of the device based on the simplest sampled grating theory. (b) SGR2 is tuned. The ten single modes span a range of 63 cm\(^{-1}\).

3.6 Conclusion

We have fabricated a three-section sampled grating QCL and provided a clear demonstration of discrete mode hopping across the entire repeat period of 26 cm\(^{-1}\)
Chapter 3: Widely tunable mid-infrared quantum cascade lasers using sampled grating reflectors

with no missing modes. Furthermore, we have demonstrated that the device is capable of lasing at additional modes not predicted by the conventional sampled grating theory, yielding a tuning range of 0.46 $\mu$m ($63 \text{ cm}^{-1}$), from 8.32 to 8.78 $\mu$m. By varying the driving conditions, the mode competition is determined by more factors than simply the reflectivity product of the mirrors, and lasing at additional modes is possible. More work is needed to determine whether these additional modes can be controllably and repeatably demonstrated across multiple devices. Additionally, continuous tuning should be demonstrated by heating both mirrors simultaneously, which can in principle also be achieved by a similar pulsing scheme with three independent electrical contacts. The discrete tuning range achieved in our device spans 5.4% of the center wavelength, which is larger than the recently demonstrated sampled grating DFB device that achieved a range of 2.39% around a center wavelength of 4.79 $\mu$m [41]. Additionally, the tuning mechanism we employ—namely, the adjustment of the durations and relative delays of pulses applied to various sections of the device to independently control the temperature in each section—is a robust tuning method for pulsed lasers; this method has an advantage over DC bias tuning because the magnitude of the driving current is the same for many different modes, resulting in a more consistent output power for all modes.
Chapter 4

Lasers with distributed loss have a sublinear output power characteristic

4.1 Abstract

It is a generally accepted fact of laser physics that in a homogeneously broadened gain medium, above threshold the output power of the laser grows linearly with the pump power. The derivation requires only a few simple lines in laser textbooks, and the linear growth is a direct result of the fact that above threshold, the intracavity optical intensity will increase to the point that the gain is saturated to the level of the net loss—so-called gain “pinning” or “clamping.” Such a derivation, however, assumes that the mirror loss is distributed—the approximation of uniform gain saturation—which is only a good assumption for cavities whose end mirrors have reflectivities close to one. Furthermore, in gain media with a distributed loss there is a maximum achievable intracavity intensity that in turn limits the output power.
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

We show that the approximation of uniform gain saturation leads to output powers that violate this limit. More specifically, for lasers with low mirror reflectivities that also have distributed loss, we prove that the output power grows sublinearly with the pump power close to threshold. Further after threshold the output grows linearly, but with a slope efficiency that can be substantially smaller than predicted by the uniform gain saturation theory, with the largest deviation occurring for traveling-wave lasers and asymmetric Fabry-Pérot lasers. These results are particularly applicable to semiconductor lasers, and specific applications to quantum cascade lasers are discussed.

4.2 Introduction

One of the signature characteristics of laser action in homogeneously broadened media is the linear growth of the output power with the pump power above the lasing threshold. In textbook derivations [45–51] this behavior is seen to be the direct result of gain clamping: above threshold, a steady state can only be reached when the intracavity optical intensity is such that the rate of stimulated emission reduces the population inversion to its threshold value, thereby always pinning the gain to the level of the net loss. Such a derivation, however, assumes that the gain saturates uniformly—in other words, that the intensity is constant throughout the gain medium—which is only a good assumption for lasers with high-reflectivity end mirrors. Surprisingly, we will show that this assumption also leads to the correct output power in the absence of distributed loss, no matter how small the reflectivities of the end mirrors. However, the uniform gain saturation approximation leads to unphysical predictions for the output power in cavities with large out-coupling
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

that also contain a distributed intrinsic loss. The key point is that an unsaturable distributed loss imposes a maximum attainable intracavity intensity (which increases with the pump power) at which the number of photons generated by the saturated gain per unit length is equal to the number of photons discarded per unit length by the loss [52–54]. The uniform gain saturation treatment does not properly account for this simple fact, causing it to overestimate the output power of the laser. We will apply the equations first popularized by Rigrod [53,55–59] to account for the nonuniformity of the gain saturation and the lumped mirror losses, and demonstrate that—in the presence of distributed loss—the output power grows sublinearly with the pump power after threshold. Further past threshold the output grows linearly, but with a slope efficiency that can be significantly less than one would expect from the uniform gain saturation theory. The deviation of the output power from the prediction of the uniform gain saturation theory is largest for traveling-wave lasers—which display the largest variation in intracavity intensity—and also very significant for highly asymmetric Fabry-Pérot (FP) cavities (i.e., one mirror with reflectivity much larger than the other). In symmetric FP cavities the deviation is the smallest, but can still be significant for particularly large values of distributed loss and out-coupling.

Much work has been done on diode lasers—the prototypical example of a laser with low facet reflectivities—to understand the nonuniformity of the gain and intracavity intensity. In this context the nonuniformity is often referred to as (long-range) spatial hole burning (SHB), but we prefer to reserve the term “SHB” to describe only the (short-range) nonuniformity due to standing-wave effects. An early theoretical work to understand the nonuniformity was given in [60], and experiments
have confirmed the expected longitudinal gain distribution [61,62]. The fact that a nonuniform intensity distribution, in the absence of distributed loss, does not affect the power output was noted in [63], and we will generalize this result. The impact of distributed loss on the slope efficiency for a few specific cases was well-illustrated both computationally and experimentally in [64], but the understanding of the effect was limited. Our goal is to provide a physical understanding of the sublinear power output, which is a simple consequence of the intracavity intensity limit imposed by the distributed loss.

In what follows, we begin by explaining the geometry of the ring and FP lasers under consideration and introduce the Rigrod equation for homogeneously broadened gain media. A physical picture is presented to understand the uniform gain approximation in which the lumped transmission losses at the mirrors are replaced with a distributed out-coupling loss. With this picture in mind, it is simple to understand why the uniform gain approximation must fail for lasers with distributed intrinsic loss. Then, taking into account nonuniform gain saturation, we first analyze the traveling-wave laser, which yields an explicit formula for the output power. We also draw general conclusions about the behavior of both symmetric and asymmetric FP lasers, and compare them to the ring. We provide simple formulas that allow one to determine when the nonuniform gain saturation treatment should be used, and to estimate the magnitude of the discrepancy between the two theories. Finally, we point out how these new results are used to optimize laser design, using the quantum cascade laser (QCL) as a model system.

A FP laser cavity is shown in Fig. 4.1(a), which comprises a gain medium.
of length $L$ placed between two lossless mirrors with reflectivities $R_1$ and $R_2$ (and transmissivities $1-R_1$ and $1-R_2$). (The facets of the gain medium itself do not reflect, either through the use of Brewster-cut facets or anti-reflection coatings.) The intensity envelopes of the right and left-propagating waves are $I^+(z)$ and $I^-(z)$. Following the notation of Rigrod [56], we will use the dimensionless ratios $\beta^+(z) = I^+(z)/I_{\text{sat}}$ and $\beta^-(z) = I^-(z)/I_{\text{sat}}$, where $I_{\text{sat}}$ is the saturation intensity of the medium. The unsaturated gain, which is typically proportional to the pump power, is $g_0$ per unit length, and the distributed intrinsic loss is $\alpha_0$ per unit length. (We take $g_0$ to be independent of position, which assumes uniform temperature and--for injection lasers--uniform current density [60,61].) In Fig. 4.1(b), the same gain medium is employed in a traveling-wave configuration by placing it within a ring cavity comprising two perfect reflectors and one out-coupling mirror of reflectivity $R$. We will consider the unidirectional mode for which only $\beta^+$ is nonzero (achieved in practice with an optical isolator).

For a homogeneously broadened gain medium, the intensity envelopes obey the equations [45,56–59,65]

$$\frac{1}{\beta^+} \frac{d\beta^+}{dz} = -\frac{1}{\beta^-} \frac{d\beta^-}{dz} = \frac{g_0}{1+\beta^+ + \beta^-} - \alpha_0. \quad (4.1)$$

We will consider only the homogeneously broadened case here; special cases of inhomogeneous broadening are discussed in [53,65,66]. By dealing only with the intensity envelope of the field, population grating effects resulting from SHB are not taken into account, but this treatment is nevertheless known to yield results in good agreement with experiment [45]. Equation 4.1 applies strictly to a single transverse and longitudinal mode; in a multimode laser the shape of the gain spectrum becomes
important as does the mutual cross-saturation of the various frequency components. While SHB is known to lead to multimode operation in homogeneously broadened lasers [65], SHB is not as significant in ring and asymmetric FP lasers, which display predominantly traveling-wave rather than standing-wave character. It is worth noting that our main conclusion in this paper is most relevant to these two types of cavities where SHB is less important. Moreover, we expect our conclusions to still be valid in the multimode regime because the essential physics behind the results—namely, the intensity limit imposed by the distributed loss—is very general.

While Eq. 4.1 (together with appropriate boundary conditions) is needed to solve for the laser output power above threshold, exactly at threshold the gain is completely unsaturated (i.e., intensity is zero) so the threshold condition is derived
simply by demanding that the roundtrip unsaturated gain compensates both the intrinsic and out-coupling losses of the cavity. (Note that this derivation applies only when spontaneous emission is very weak, which is a good assumption for many kinds of lasers. A treatment that includes gain saturation at threshold due to spontaneous emission is given in [67, 68].) To facilitate comparisons between the FP and ring lasers, we choose the reflectivity of the ring out-coupler to be $R = \sqrt{R_1 R_2}$ so that both cavities have the same threshold, given by

$$g_{0,\text{th}} = \alpha_0 + \frac{1}{L} \ln \left( \frac{1}{R} \right).$$

(4.2)

### 4.3 Distributed Loss Approximation

The second term on the right-hand side of Eq. 4.2 is commonly called the mirror loss $\alpha_m = \ln(1/R)/L$, so that the total loss can be written as $\alpha_0 + \alpha_m$. We emphasize, however, that these two contributions to the loss behave very differently, $\alpha_0$ being a distributed loss and $\alpha_m$ a lumped loss. A significant approximation made in textbook derivations of the laser output power [50, 51]—though not always explicitly stated—is to treat $\alpha_m$ as a distributed loss. In this way, all losses of the cavity become distributed, so we will refer to this as the distributed loss approximation (DLA). One way to think of the DLA is to imagine that the end mirrors of the cavity are replaced by perfect reflectors, and light escapes instead from any point within the gain medium with decay constant $\alpha_m$. (Equivalently, a photon escape time $\tau_{\text{esc}}$ is often defined as $\tau_{\text{esc}}^{-1} \equiv \alpha_m v_g$, where $v_g$ is the group velocity of light in the cavity.) Because the end mirrors have unity reflectivity, in the steady state the intracavity intensity does not vary with position; the saturated gain is exactly cancelled by the sum of the intrinsic loss and
mirror loss at every position in the cavity. The DLA is therefore more commonly referred to as the uniform gain saturation approximation. Mathematically, Eq. 4.1 becomes

\[
\frac{1}{\beta^+} \frac{d\beta^+}{dz} = \frac{g_0}{1 + \beta} - (\alpha_0 + \alpha_m) = 0,
\]

(4.3)

where \( \beta = \beta^+ + \beta^- \) is the total intensity. Equation 4.3 is easily solved for \( \beta \) and multiplied by the distributed out-coupling \( \alpha_m L \) (as opposed to the lumped out-coupling \( 1 - R \), which is not the relevant out-coupling in the DLA) to get the total out-coupled intensity

\[
\beta_{\text{DLA}}^\text{out} = \frac{\alpha_m L}{\alpha_0 + \alpha_m} (g_0 - g_{0,\text{th}}).
\]

(4.4)

Equation 4.4 is the well-known formula for the output power of a laser. In the DLA, both the FP and ring lasers emit the same intensity \( \beta_{\text{DLA}}^\text{out} \). Note that the linearity of \( \beta_{\text{DLA}}^\text{out} \) with the unsaturated gain \( g_0 \) and therefore, in most cases, the pump power—follows from the DLA almost immediately.

There is a critical shortcoming of the DLA that is easy to understand. In a gain medium with intrinsic loss, whether an amplifier or a laser, there is a maximum attainable intracavity intensity \( \beta^\text{max} \) at which the number of photons generated by stimulated emission per unit length is exactly equal to the number of photons dissipated by the intrinsic loss per unit length [52–54]. By setting Eq. 4.1 to zero, we find that the total steady-state intracavity intensity, \( \beta^+ + \beta^- \), cannot exceed

\[
\beta^\text{max} = \frac{1}{\alpha_0} (g_0 - \alpha_0)
\]

(4.5)
at any position. This, in turn, places an upper limit on the out-coupled intensity. For simplicity, we consider the ring laser to explain the basic point: the output cannot
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

Figure 4.2: (a) A sketch illustrates that if the slope of $\beta_{\text{DLA}}^{\text{out}}$ (dashed) exceeds the slope of $\beta_{\text{ring}}^{\text{out, max}}$ (solid), the DLA prediction exceeds the allowable maximum output of the ring laser above a certain gain. Therefore, we can expect the actual output $\beta_{\text{ring}}^{\text{out}}$ to qualitatively follow the dotted curve. Inset: the parameter ranges of $\alpha_0 L$ and $R$ that result in such a violation, namely, those which satisfy Inequality 4.7, are shaded in light blue.

exceed the mirror transmission times $\beta^{\text{max}}$,

$$
\beta_{\text{ring}}^{\text{out, max}} = (1 - R)\beta^{\text{max}} = \frac{1 - R}{\alpha_0} (g_0 - \alpha_0). 
$$

(Similar formulas apply to the FP laser in Eqs. 4.22 and 4.23.) It turns out that the DLA, by way of replacing the lumped mirror transmission $1 - R$ with the distributed out-coupling $\alpha_m L$, yields predictions for the output intensity that violate this maximum under certain conditions. Specifically, whenever the slope efficiency of $\beta_{\text{DLA}}^{\text{out}}$ is greater than the slope of $\beta_{\text{ring}}^{\text{out, max}}$,

$$
\frac{\alpha_m L}{\alpha_0 + \alpha_m} > \frac{1 - R}{\alpha_0}, 
$$

the sketch in Fig. 4.2 illustrates that $\beta_{\text{DLA}}^{\text{out}}$ will exceed $\beta_{\text{ring}}^{\text{out, max}}$ above some value for $g_0$. In the next section we will solve explicitly for the output $\beta_{\text{ring}}^{\text{out}}$ taking into account nonuniform gain saturation, but we can already guess that the curve should
Lasers with distributed loss have a sublinear output power characteristic. It resembles the dotted line in Fig. 4.2: it has the same threshold as the DLA predicts, but the output power must eventually be limited by $\beta_{\text{ring}}^{\text{out,max}}$. The parameter ranges of $\alpha_0 L$ and $R$ that satisfy Inequality 4.7 are shown in light blue in the inset; in general, as $\alpha_0 L$ becomes larger and $R$ smaller, the limiting influence of $\beta_{\text{max}}$ becomes more relevant and nonuniform gain saturation must be accounted for. For example, for semiconductor lasers ($R \approx 0.3$), Fig. 4.2 indicates that the DLA is certainly problematic when $\alpha_0 L$ exceeds 1.7, and we will show in subsequent sections that the discrepancy is measurable even for smaller values of $\alpha_0 L$.

### 4.4 Nonuniform gain saturation

We now address the behavior of the laser when the lumped nature of the mirror losses is properly accounted for, and the gain saturates nonuniformly within the cavity. The general solution of Eq. 4.1 is given in [56], which we will restate here. It can be deduced from Eq. 4.1 that the product $\beta^+(z)\beta^-(z)$ is independent of $z$, so we designate the quantity $\beta_0^2 \equiv \beta^+(z)\beta^-(z)$. By making the substitution $\beta^-(z) = \beta_0^2/\beta^+(z)$, Eq. 4.1 is expressed entirely in terms of $\beta^+(z)$, which can then be integrated to yield

$$
\alpha_0 L - \ln(\beta_1^+ / \beta_2^+) = \frac{g_0 \ln[F(\beta_1^+)/F(\beta_2^+)][F(\beta_1^+)/F(\beta_2^+)]}{\sqrt{(g_0 - \alpha_0)^2 - (2\alpha_0 \beta_0)^2}},
$$

where

$$F(\beta_i^+) \equiv \frac{\sqrt{(g_0 - \alpha_0)^2 - (2\alpha_0 \beta_0)^2} + (g_0 - \alpha_0 - 2\alpha_0 \beta_i^+) + (g_0 - \alpha_0 - 2\alpha_0 \beta_i^+)}{\sqrt{(g_0 - \alpha_0)^2 - (2\alpha_0 \beta_0)^2} - (g_0 - \alpha_0 - 2\alpha_0 \beta_i^+)}. \quad (4.9)
$$

The subscript $i$ on $\beta^+$ is either 1 or 2, and denotes evaluation of the intensity at the left facet ($z = 0$) or right facet ($z = L$), respectively. Note that Eqs. 4.8 and 4.9 apply to both the FP and ring lasers. For the FP, we use the boundary conditions at
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

the mirrors

\[ \beta_1^+ = R_1 \beta_1^-, \quad \beta_2^- = R_2 \beta_2^+ \]  \hspace{1cm} (4.10)

together with

\[ \beta_0^2 = \beta_1^+ \beta_1^- = \beta_2^+ \beta_2^- \]  \hspace{1cm} (4.11)

to show that \( \beta_1^+ = R \beta_2^+ \) and \( \beta_0 = \sqrt{R_2} \beta_2^+ \), which allows us to express Eq. 4.8 entirely in terms of \( \beta_2^+ \). Still, in the general FP case Eq. 4.8 remains an implicit formula for \( \beta_2^+ \) which must be solved numerically. For the ring laser we substitute \( \beta_1^+ = R \beta_2^+ \) and \( \beta_0 = 0 \) into Eq. 4.8, which yields an analytic solution that will be discussed shortly.

Before we investigate Eq. 4.8 for the ring cavity and FP cavity in detail, we note two limiting cases for which the total output intensity as predicted by the nonuniform gain saturation theory in Eq. 4.8 reduces to that of the DLA: 1) small cavity out-coupling and arbitrary intrinsic loss \( \alpha_0 \) and 2) arbitrary out-coupling and \( \alpha_0 = 0 \) (a surprising result, since the intensity variation in the cavity can be very large yet the DLA still predicts the correct output power [63]).

When the reflectivity of both mirrors approaches one and there is also non-zero distributed loss (so that \( 1 - R \ll \alpha_0 L \)), Eqs. 4.8 and 4.9 can be Taylor expanded to first order in the parameters \( 1 - R_1 \) and \( 1 - R_2 \), which yields the total output power

\[ \beta_{\text{out}} \approx \frac{1 - R}{\alpha_0} (g_0 - \alpha_0). \]  \hspace{1cm} (4.12)

The DLA yields the same result as Eq. 4.12 when \( \beta_{\text{out}}^{\text{DLA}} \) in Eq. 4.4 is expanded in the same limits. The agreement of the two theories in this limit is due to the highly uniform intracavity intensity (as a result of the high mirror reflectivities) together with the fact that \( \ln(1/R) \approx 1 - R \), so that the distributed out-coupling of the DLA is not a bad approximation of the true lumped nature of the mirror losses.
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

When $\alpha_0 = 0$ there is no finite limiting value of the intracavity intensity $\beta^{\text{max}}$; still, this is not enough of a reason to expect the DLA to be a good approximation because there can be significant intensity variation in the cavity—particularly for small $R$—that would seem to necessitate the nonuniform gain saturation treatment. Taking the limit of Eqs. 4.8 and 4.9 as $\alpha_0$ goes to 0 yields

$$g_0L = \ln \left( \frac{\beta_2^+}{\beta_1^+} \right) + \beta_2^+ - \beta_1^+ - \beta_0^2 \left( \frac{1}{\beta_2^+} - \frac{1}{\beta_1^+} \right). \tag{4.13}$$

For the FP laser, this can be solved for the intensity emitted from each facet

$$\beta_1^{\text{out}} = (1 - R_1) \beta_1^- = \frac{(1 - R_1)L(g_0 - g_{0,th})}{1 - R_1 + \sqrt{R_1/R_2 - \sqrt{R_1R_2}}}, \tag{4.14}$$

$$\beta_2^{\text{out}} = (1 - R_2) \beta_2^- = \frac{(1 - R_2)L(g_0 - g_{0,th})}{1 - R_2 + \sqrt{R_2/R_1 - \sqrt{R_1R_2}}}, \tag{4.15}$$

and the total emitted intensity

$$\beta^{\text{out}} = \beta_1^{\text{out}} + \beta_2^{\text{out}} = L(g_0 - g_{0,th}). \tag{4.16}$$

For the ring laser, it can be shown that the total emitted intensity is also given by Eq. 4.16 by setting $\beta_0 = 0$ in Eq. 4.13. Interestingly, the DLA predicts exactly the same total output power (seen by setting $\alpha_0 = 0$ in Eq. 4.4) as the nonuniform gain saturation treatment in this limit. (Although, the two theories do not agree on the fraction of the total power coupled out of each mirror of the FP laser. This discrepancy is, however, small in most cases. In particular, there is no discrepancy between the theories for the case of $R_1 = R_2$—half of the power leaves from each mirror—or when one of the mirrors has unity reflectivity—all of the power is coupled out of the other mirror.) As far as I can tell, there is no obvious reason for the DLA to yield the same total output power as the nonuniform treatment when $\alpha_0 = 0$. If nonradiative
recombination is neglected [63], then the result can be explained by conservation of energy because the only pathway for electrons to descend from the excited to ground state is through stimulated emission. In other words, the number of photons emitted by the cavity depends only on the total number of electrons passing through the cavity, regardless of the intracavity intensity distribution [63]. Our result is more surprising because it allows electrons to recombine nonradiatively, so this argument does not apply. (Note: to fully explain these statements would require a derivation of the Rigrod equation from rate equations, which is not difficult but outside the scope of our work. For the interested reader, we also point out that we use a linear relationship between \( g_0 \) and carrier density whereas [63] uses a logarithmic relationship.)

From here on, we will focus on the case of large out-coupling and large intrinsic loss, for which the predictions of the nonuniform gain saturation theory deviate significantly from those of the DLA.

### 4.4.1 Ring laser

Unlike the FP, the equations for the unidirectional ring laser can be solved for an explicit expression of the output power. (The existence of this solution--but not its expression--was stated in [56].) Starting from Eqs. 4.8 and 4.9 and using \( \beta_0 = 0 \) and \( \beta_1^+ = R\beta_2^+ \), we find

\[
\beta_{\text{out}} = (1 - R) \left( g_0 / \alpha_0 - 1 \right)
\times \frac{R \exp[(1 - \alpha_0/g_0)(\alpha_0 + \alpha_m)L] - 1}{R \exp[(1 - \alpha_0/g_0)(\alpha_0 + \alpha_m)L] - R}
\]

(4.17)

We will explore various limits of Eq. 4.17. Close to threshold, the slope efficiency can be found by Taylor expanding the derivative of Eq. 4.17 with respect to \( g_0 \) to first
order in the parameter \( (g_0 - g_{0,th})/(\alpha_0 + \alpha_m) \), which gives

\[
\frac{d\beta_{\text{ring}}^\text{out}}{dg_0} = \frac{\alpha_m L}{\alpha_0 + \alpha_m} - \frac{\alpha_0 L}{(\alpha_0 + \alpha_m)^2} \left( 1 + \frac{R}{1 - R} \ln\left(\frac{1}{R}\right) - 2 \right) (g_0 - g_{0,th}).
\]

(4.18)

(We have also only retained terms to first order in \( \alpha_0 L \) for simplicity, though the following conclusions hold for large \( \alpha_0 L \).) At threshold (i.e., \( g_0 = g_{0,th} \)), the slope efficiency is given by the first term of Eq. 4.18, which is the same as the prediction of the DLA in Eq. 4.4. However, while the DLA predicts a constant slope efficiency above threshold, in fact the slope efficiency decreases as \( g_0 \) increases past \( g_{0,th} \) as given by the second term of Eq. 4.18. (Note that the term in large parentheses is always positive for \( 0 < R < 1 \).)

As \( g_0 \) continues to increase, the slope efficiency asymptotically approaches a constant value, which we can see by Taylor expanding Eq. 4.17 in the high-gain limit \( g_0 \gg g_{0,th} \) (without making any assumption about \( \alpha_0 L \)), which gives

\[
\beta_{\text{ring}}^\text{out} \big|_{g \gg g_{0,th}} = \frac{(1 - R)(1 - e^{-\alpha_0 L})}{\alpha_0 (1 - Re^{-\alpha_0 L})} (g_0 - g_{0,th}^{\text{eff}})
\]

where

\[
g_{0,th}^{\text{eff}} \equiv (\alpha_0 + \alpha_m)(\alpha_0 L) \left( \frac{1}{1 - e^{-\alpha_0 L}} - \frac{1}{1 - Re^{-\alpha_0 L}} \right) + \alpha_0
\]

(4.19)

is an “effective threshold” that would be found by linearly extrapolating the power output at high gain back to the \( g_0 \) axis. We will refer to the derivative of Eq. 4.19 with respect to \( g_0 \) as the “asymptotic slope efficiency,” which is clearly independent of \( g_0 \). In the limit of “moderate to large” intrinsic loss, \( \exp(\alpha_0 L) \gg 1 \), (and arbitrary \( R \) ) Eq. 4.19 reduces to

\[
\beta_{\text{ring}}^\text{out} \big|_{g \gg g_{0,th}, \exp(\alpha_0 L) \gg 1} = \frac{1 - R}{\alpha_0} (g_0 - \alpha_0).
\]

(4.21)
which is recognizable as $\beta_{\text{ring}}^{\text{out, max}}$ of Eq. 4.6. Thus, we see that for moderate to large $\alpha_0 L$, at high gain the intracavity intensity reaches its limiting value $\beta_{\text{max}}$, so the output intensity follows $\beta_{\text{ring}}^{\text{out, max}}$, as illustrated by the dotted line in Fig. 4.2. For smaller $\alpha_0 L$ that does not satisfy $\exp(\alpha_0 L) \gg 1$, the output intensity requires the full Eq. 4.19; in this case the intracavity intensity never grows sufficiently to approach $\beta_{\text{max}}$, but the asymptotic slope efficiency is nevertheless smaller than the DLA slope efficiency.

To understand quantitatively how much the asymptotic slope efficiency differs from the near-threshold slope efficiency, let us look at the ratio of the slope efficiency in the high-gain limit (slope of Eq. 4.19) to the slope efficiency at threshold (the first term of Eq. 4.18, which is the same as the slope efficiency predicted by the DLA). This ratio is plotted in Fig. 4.3 as a function of $R$ for five different values of the product $\alpha_0 L$. As expected, the discrepancy between the asymptotic and threshold slope efficiency

Figure 4.3: Ring laser. The ratio of the asymptotic to the threshold slope efficiency as a function of $R$ for various values of $\alpha_0 L$. 

71
is greatest for small $R$ and large $\alpha_0 L$. The effect of $\alpha_0 L$ can be understood with recourse to Fig. 4.2: as $\alpha_0$ increases the slope of $\beta_{\text{ring}}^{\text{out}, \max}$ decreases, and as $L$ increases the slope of $\beta_{\text{DLA}}^{\text{out}}$ increases. In either case, an increase in the product $\alpha_0 L$ widens the disparity between the asymptotic and threshold slope efficiency.

![Figure 4.4](image_url)

Figure 4.4: The value $g_0^{\text{asmp}}/g_{0,\text{th}}$—the gain at which the power output reaches 95% of the asymptotic solution—is plotted for values of (a) $\alpha_0 L \leq 2$ and (b) $\alpha_0 L \geq 3$, which shows that $g_0^{\text{asmp}}$ decreases monotonically with increasing $\alpha_0 L$ beyond $\alpha_0 L = 3$.

Finally, while the asymptotic output intensity was derived in the limit $g_0 \gg g_{0,\text{th}}$, it turns out that this is not a very stringent limit. In fact, the full solution in Eq. 4.17 closely approaches the asymptotic solution in Eq. 4.19 when $g_0$ is only a few multiples of $g_{0,\text{th}}$, and often when $g_0$ is even smaller. (This is significant because in some lasers, such as QCLs, $g_0$ can never exceed more than a few multiples of $g_{0,\text{th}}$, and so this asymptotic regime is still experimentally accessible.) To justify this statement, we look at the ratio of the output power given by the exact expression of Eq. 4.17 to that of the asymptotic expression of Eq. 4.19, $\beta_{\text{out}}/(\beta_{\text{out}}|_{g_0 \gg g_{0,\text{th}}})$, and define $g_0^{\text{asmp}}$ as the gain above which this ratio is greater than 0.95. (Our choice of 0.95 is arbitrary, but is a good indicator for when the exact solution approaches
the asymptotic solution.) In Figs. 4.4(a) and 4.4(b), \( g_0^{\text{asmp}}/g_{0,\text{th}} \) is plotted against \( R \) for several different values of \( \alpha_0L \). The most striking conclusion is that provided \( R > 0.01 \) (a condition satisfied by nearly all practical lasers), the asymptotic theory becomes very accurate for \( g_0 > 2.6g_{0,\text{th}} \). In Fig. 4.4(a), we see that as \( \alpha_0L \) increases from 0.1 to 2 at fixed \( R \), \( g_0^{\text{asmp}}/g_{0,\text{th}} \) increases as well. However, for \( \alpha_0L \geq 3 \) as shown in Fig. 4.4(b), this behavior changes, and \( g_0^{\text{asmp}}/g_{0,\text{th}} \) instead decreases monotonically with increasing \( \alpha_0L \) at fixed \( R \). Combining the information in Figs. 4.3 and 4.4(b), we reach an important conclusion. As \( \alpha_0L \) increases beyond \( \approx 3 \), two things happen: 1) the asymptotic slope efficiency becomes significantly smaller than the threshold slope efficiency, and 2) the asymptotic regime is reached even closer to threshold. This second trend can be understood easily with recourse to Fig. 4.2: note that the gain-intercept of \( \beta_{\text{ring}}^{\text{out,max}} \) is \( \alpha_0 \) while the laser threshold is \( \alpha_0 + \alpha_m \). As \( \alpha_0L \) increases at fixed \( R \), the intercept \( \alpha_0 \) approaches \( \alpha_0 + \alpha_m \), and so the laser must enter the asymptotic regime sooner after threshold. Let us consider one numerical example: for \( \alpha_0L = 10 \) and \( R = 0.25 \), the asymptotic slope efficiency is 62\% of the threshold slope efficiency (see Fig. 4.3), and the asymptotic regime is reached once \( g_0 \geq 1.4g_{0,\text{th}} \).

This example will be extended to the FP laser in the next section.

### 4.4.2 FP laser

The maximum output intensity of the ring laser \( \beta_{\text{ring}}^{\text{out,max}} \) in Eq. 4.6 was derived by assuming that the intracavity intensity reaches \( \beta^{\text{max}} \) at the out-coupling mirror. Similarly, the maximum output intensities of the symmetric and maximally asym-
metric FP lasers are

\[ \beta_{sFP}^{\text{out, max}} = 2 \left( \frac{1 - R}{1 + R} \right) \beta_{sFP}^{\text{max}}, \]  
\[ \beta_{aFP}^{\text{out, max}} = \left( \frac{1 - R^2}{1 + R^2} \right) \beta_{aFP}^{\text{max}}. \]  

(For the maximally asymmetric FP, the reflectivity of the left mirror is unity and the other is \( R_2 = R^2 \).) The denominators \( 1 + R \) and \( 1 + R^2 \) in Eqs. 4.22 and 4.23, respectively, arise from the need to transform the total intensity \( \beta_{sFP}^{\text{max}} \) into the right-traveling component \( \beta_{sFP}^{\text{out, max}} \), which is then multiplied by the transmissivity \( 1 - R \) (and a factor of 2 for the two mirrors) or \( 1 - R^2 \), respectively. Inequalities similar to Inequality 4.7 can easily be written down using Eqs. 4.22-4.23, which then serve as useful indicators for when to be wary of the DLA solution for FP lasers.

![Figure 4.5](image_url)

Figure 4.5: (a) \( \beta_{sFP}^{\text{out}} \) vs. \( g_0 \) for \( \alpha_0 = 20 \text{ cm}^{-1}, L = 5 \text{ mm}, \) and \( R = 0.25 \), shown for the DLA (black), symmetric FP with \( R_1 = R_2 = 0.25 \) (red), maximally asymmetric FP with \( R_1 = 1 \) and \( R_2 = 0.25^2 = 0.0625 \) (blue), and ring laser with \( R = 0.25 \) (green). The slope efficiencies of the curves (i.e., first derivatives) are shown in the inset, together with the corresponding slopes of \( \beta_{sFP}^{\text{out, max}} \) for each laser. (b) The intracavity intensity at the right facet, \( \beta_{sFP}^{\text{out, max}} + \beta_{aFP}^{\text{out, max}} \), is normalized to \( \beta_{sFP}^{\text{max}} \) and plotted against \( g_0 \) over a larger range than in (a).
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

Figure 4.6: For the same parameters as those used in Fig. 4.5 \((\alpha_0 = 20 \text{ cm}^{-1}, \ L = 5 \text{ mm}, \ R = 0.25)\), the intensity envelopes \(\beta^+\) and \(\beta^-\) of the right and left-propagating waves are plotted for \(1.2g_{0,\text{th}}\) to \(3g_{0,\text{th}}\) (in steps of \(0.3g_{0,\text{th}}\)) for a (a) symmetric FP with \(R_1 = R_2 = 0.25\), (b) maximally asymmetric FP with \(R_1 = 1\) and \(R_2 = 0.25^2 = 0.0625\), and (c) ring laser with \(R = 0.25\). The total intracavity intensity, \(\beta^+ + \beta^-\), is normalized to \(\beta^{\text{max}}\) and plotted in (d),(e), and (f) for the same three lasers, respectively.

Since we cannot analytically examine the output power of the FP laser as we did for the ring laser, we consider a concrete example for which \(\alpha_0 = 20 \text{ cm}^{-1}\), \(L = 0.5 \text{ cm}\), and \(R = 0.25\), and solve numerically for the output power. (These are representative values for a QCL with emission wavelength between 8 and 12 \(\mu\text{m}\) [69].)

In Fig. 4.5, the total output intensity \(\beta^{\text{out}}\) as a function of \(g_0\) is shown for three lasers: 1) symmetric FP with \(R_1 = R_2 = 0.25\), 2) maximally asymmetric FP with \(R_1 = 1\), \(R_2 = 0.25^2 = 0.0625\), and 3) ring laser with \(R = 0.25\). According to the DLA, all three lasers should have the same output intensity, and this curve is plotted as well. The maximum output intensity for each laser, given by Eqs. 4.6 and 4.22-4.23, is plotted in dashed lines. In the inset, the slope efficiency of each laser is shown.
alongside the derivative of $\beta^{\text{out, max}}$; for every laser, the slope of $\beta^{\text{out, max}}$ is less than the slope efficiency of the DLA, which means the curve $\beta^{\text{out, max}}$ will fall beneath $\beta^{\text{out}}_{\text{DLA}}$ above some value of $g_0$. This is clearly seen in Fig. 4.5 for the ring and asymmetric FP, and occurs outside the plotted domain (at $g_0 = 212$ cm$^{-1}$) for the symmetric FP. We see that all three lasers have the same slope efficiency at threshold, and the value is correctly predicted by the DLA. (For the ring laser we already proved this fact in the previous section.) As the gain increases, the slope efficiencies of all three lasers decrease, as they must because once the asymptotic regime is reached the slope efficiency must be less than the slope of $\beta^{\text{out, max}}$. In the asymptotic regime, the symmetric FP has the largest slope efficiency (96% of threshold value), followed by the asymmetric FP (72% of threshold value) and then the ring (62% of threshold value). In this example, the DLA is not a bad approximation for the symmetric FP because $\beta^{\text{out, max}}_{\text{sFP}}$ does not violate $\beta^{\text{out}}_{\text{DLA}}$ over a large range of gain. However, the DLA does not provide accurate output intensities for the asymmetric FP and ring, for which the limiting effects of $\beta^{\text{out, max}}_{\text{aFP}}$ and $\beta^{\text{out, max}}_{\text{ring}}$ are stronger and occur soon after threshold.

Another way to understand the output power of the lasers in Fig. 4.5(a) is shown in Fig. 4.5(b): here, the intracavity intensity at the right mirror (where it is a maximum), $\beta^+_2 + \beta^-_2$, is normalized to $\beta^{\text{max}}$ and plotted as a function of $g_0$. Note that the normalization factor $\beta^{\text{max}}$ is itself an increasing function of $g_0$. In the ring laser and asymmetric FP, shortly after threshold the total intensity at the right mirror quickly approaches $\beta^{\text{max}}$. (To be exact, in the asymptotic regime the intracavity intensities approach $0.99997\beta^{\text{max}}$ and $0.99987\beta^{\text{max}}$ in the ring and asymmetric FP, respectively.)
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

In the symmetric FP, the total intensity at the right mirror asymptotically approaches the slightly smaller value of \(0.974 \beta_{\text{max}}\), although for larger values of \(\alpha_0 L\) it can be shown to also approach \(\beta_{\text{max}}\). Thus, for sufficiently large \(\alpha_0 L\), Eqs. 4.6 and 4.22-4.23 do not provide merely an upper limit, but rather a good approximation of the output intensity in the asymptotic regime. What do we mean by “sufficiently large \(\alpha_0 L\)?”

For the ring laser we previously showed the necessary condition to be \(\exp(\alpha_0 L) \gg 1\), but for the FP this is not easy to answer without numerical calculation. As a general trend, however, the ring laser always has the largest intracavity intensity at the out-coupling mirror, followed by the asymmetric FP and then the symmetric FP. Thus, the symmetric FP requires the largest amount of loss \(\alpha_0 L\) in order for the intensity to reach \(\beta_{\text{max}}\), followed by the asymmetric FP and then the ring. To explain this, it helps to plot the intracavity intensity along the entire length of the cavity, as shown in Fig. 4.6 for the same laser parameters as used in Fig. 4.5. In the symmetric FP, the left and right-propagating waves have the most similar intensities at each position; the cross-saturation of the two waves hinders the growth of each and prevents the total intracavity intensity from reaching \(\beta_{\text{max}}\). In the asymmetric FP, \(\beta^+\) (which travels from the high-reflectivity mirror to the lower-reflectivity mirror) is much more intense than \(\beta^-\). Therefore, \(\beta^-\) does not cross-saturate \(\beta^+\) significantly, which allows \(\beta^+\) to grow quickly. As a result, we see that the total intensity closely approaches \(\beta_{\text{max}}\).

In the ring laser, the absence of \(\beta^-\) allows \(\beta^+\) to grow even more quickly at small \(z\) and approach the limit \(\beta_{\text{max}}\) well before it reaches the right mirror. The key point is that cross-saturation in the symmetric FP reduces the total intracavity intensity. However, for large enough \(\alpha_0 L\), even this mechanism does not prevent the intensity
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

from reaching \( \beta^{\text{max}} \).

We can also provide a more mathematical description of the variation of the total intensity \( \beta^+ + \beta^- \) within the laser cavity. For the ring laser this is quite simple since \( \beta^- = 0 \), so the maximum and minimum occur at the facets, and their ratio is given by

\[
\left. \frac{(\beta^+ + \beta^-)_{\text{min}}}{(\beta^+ + \beta^-)_{\text{max}}}_\text{ring} \right| = R. \tag{4.24}
\]

For the FP laser, recall that \( \beta^+ \beta^- = \beta_0^2 \), so the total intensity can be written as

\[
\beta^+ + \beta^- = \beta^+ + \beta_0^2/\beta^+. \tag{4.25}
\]

The minimum total intensity is found by differentiating Eq. 4.25 and occurs at the position where the two counter-propagating waves have equal intensity, i.e., \( \beta^+ = \beta^- = \beta_0 \), and therefore \( (\beta^+ + \beta^-)_{\text{min}} = 2\beta_0 \). The maximum intensity can be found from Eq. 4.25 to occur at the facet of lower reflectivity, which we take to be \( R_2 \), and is given by \( (\beta^+ + \beta^-)_{\text{max}} = \beta_0(1 + R_2)/\sqrt{R_2} \). Thus, the total intensity variation in the FP laser is given by

\[
\left. \frac{(\beta^+ + \beta^-)_{\text{min}}}{(\beta^+ + \beta^-)_{\text{max}}} \right|_{\text{FP}} = \frac{2\sqrt{R_2}}{1 + R_2}, \tag{4.26}
\]

where \( R_2 \) is the facet of lower reflectivity. It is interesting to note that this variation in total intensity is independent of \( g_0 \) or \( \alpha_0 \)—it is determined solely by the facet reflectivities. A ratio closer to 1 indicates a more uniform intensity in the cavity. Equation 4.26 is plotted in [55] (but was derived here under more general considerations). By comparing Eqs. 4.24 and 4.26, one finds that the intensity in the FP is always more uniform than in the ring, and in most cases substantially so. Physically, the reason
for this is of course that the intensity in the FP laser is shared between two counter-
propagating waves which grow in opposite directions and strongly cross-saturate each
other.

For fixed $R = \sqrt{R_1 R_2}$, the most uniform distribution is achieved when $R_1 = R_2$. As $R_1$ increases and $R_2$ decreases, the intensity becomes more nonuniform. The least uniform distribution occurs for $R_1 = 1$ and the minimum value of $R_2 = R^2$. This is the closest that the FP laser intensity distribution can come to resembling the ring laser intensity distribution—for a given $R$—because the right-propagating intensity $\beta^+$ is as large as possible relative to the left-propagating $\beta^-$. In this case, the ratio of Eq. 4.24 to Eq. 4.26 is $(1 + R_2)/2$, which tells us that for small $R_2$ the intensity within the FP is still about twice as uniform as that of the ring.

### 4.5 Other Applications

The DLA overestimates the increase in output power that can be achieved by increasing the length of the laser. As a practical example, we consider an asymmetric FP QCL with $\alpha_0 = 15 \text{ cm}^{-1}$, high-reflectivity (HR) coating on one facet $R_1 = 1$, and low-reflectivity (LR) coating on the other $R_2 = 0.01$ (an achievable value with current mid-infrared coating technology) to increase the single-ended output power. In practice every laser has a maximum achievable gain $g_0$, so suppose our goal is to maximize the output power at $g_0 = 40 \text{ cm}^{-1}$. In Fig. 4.7, the output power $\beta^{\text{out}}$ vs. $g_0$ is plotted for two such lasers, one with $L = 4 \text{ mm}$ and one with $L = 5 \text{ mm}$, for both the DLA and nonuniform gain saturation theory. In the DLA, the longer laser outputs 12% more power at $g_0 = 40 \text{ cm}^{-1}$ ($\beta^{\text{out}} = 2.40$ instead of 2.13). In reality, however,
the longer laser can be expected to output only 3% more power ($\beta^{\text{out}} = 1.59$ instead of 1.54). The reason for this small increase is that the 4-mm laser is already emitting close to the maximum possible intensity $I_{\text{asym}}^{\text{max}}$ of Eq. 4.23 (also plotted in Fig. 4.7), so increasing the length of the laser will not increase the output significantly. In this example, a 25% increase in the length of the cavity results in only a 3% increase in power output, which clearly leads to a significant decrease in the wall-plug efficiency. More generally, for lasers with large intrinsic loss there is an optimum length beyond which the wall-plug efficiency will decrease.

The output intensity of a laser is maximized for a particular end mirror reflectivity, and this optimal reflectivity will be significantly miscalculated when using the DLA. (Note: Rigrod provides analytic formulas that account for nonuniform gain saturation to approximate the optimal reflectivity [56], but the formulas become less accurate for lasers with large intracavity intensity variation. This is the case for the asymmetric FP and ring, so here we will calculate the optimum numerically.) As before, let us consider $\alpha_0 = 15 \, \text{cm}^{-1}$, $R_1 = 1$, and $L = 4 \, \text{mm}$, though now our goal is to find the optimum reflectivity $R_2$ that maximizes the power output at $g_0 = 40 \, \text{cm}^{-1}$. The nonuniform gain saturation theory predicts an optimum reflectivity $R_2 = 5.4 \times 10^{-3}$ leading to $\beta^{\text{out}} = 1.54$. The DLA predicts an optimum reflectivity $R_2 = 5.0 \times 10^{-4}$, nearly an order of magnitude smaller. Using a coating with the DLA optimum reflectivity would result in $\beta^{\text{out}} = 1.50$, which is only a 2.6% reduction in output power relative to the true optimum. This is a small difference, but the importance of this example is that the optimal coating reflectivity is an order of magnitude larger than predicted by the DLA. This is important for active research.
in LR mid-infrared coatings: for the purpose of maximizing laser output power, we do not need to aim for reflectivities as low as we had thought.

![Graph showing output power vs. gain](image)

**Figure 4.7:** The output power $\beta_{\text{out}}$ of the maximally asymmetric FP laser with $\alpha_0 = 15 \text{ cm}^{-1}$, $R_1 = 1$, and $R_2 = 0.01$, for cavities of length 4 mm (red) and 5 mm (blue), as computed numerically accounting for nonuniform gain saturation (red and blue solid lines) and as predicted by the DLA (dashed lines). The maximum output power $\beta_{\text{out,max}}$ is also plotted (black).

One final application concerns laser characterization. The ratio of the experimentally measured slope efficiencies of two lasers—identical except for their facet coatings, whose reflectivities are known—can be used to determine the value of $\alpha_0$ by comparison with the DLA slope efficiency formula [44]. In such cases, the result can be skewed if one of the lasers is highly asymmetric, and a preferable method to determine the distributed loss would be to vary the cavity length $L$ while maintaining the cavity symmetry $R_1 = R_2$ (or use the threshold values rather than the slope efficiencies, which are unaffected by the degree of symmetry of the cavity). (Note: the results of [44] are not affected by this phenomenon because in their case one of the lasers had a HR facet and a bare facet, as opposed to a LR facet. Based on numerical
Chapter 4: Lasers with distributed loss have a sublinear output power characteristic

calculations we have done, this would not have been enough asymmetry to affect their results by an experimentally significant margin.)

4.6 Conclusion

The uniform gain saturation treatment of a laser effectively replaces the transmission of light at the end mirrors with an out-coupling mechanism distributed throughout the cavity. We referred to this as the distributed loss approximation (DLA), and showed that it is a good approximation to the nonuniform gain saturation treatment when the out-coupling is small or when the intrinsic loss of the gain medium is zero. When the intrinsic loss is not zero, there is a maximum attainable intracavity intensity $\beta^{\text{max}}$ at which the number of photons generated by stimulated emission per unit length is equal to the number of photons absorbed or scattered by the intrinsic loss per unit length. When the intensity in the laser nears this maximum, the DLA overestimates the power that can be coupled out of the cavity. The result is that the output power grows sublinearly with the pump power, in stark contrast to the famous linear relationship derived in the uniform gain saturation approximation.

We provided simple formulas in terms of $\alpha_0L$ and $R$—applicable to all FP lasers—to help determine the magnitude of the output power discrepancy between the DLA and the nonuniform treatment. The deviation is greatest for traveling-wave lasers and highly asymmetric FP lasers.

While our derivation explicitly treats only single-mode lasers and neglects SHB, in fact we expect the sublinear power output to be a quite general effect. To be completely rigorous, the standing-wave effects of SHB should be accounted for, and
in a multimode laser the gain cross-saturation of each frequency on the others complicates the mathematics significantly. Therefore, the value of $\beta_{\text{max}}$ for each mode will be different and difficult to calculate. Nevertheless, the limiting effect of $\beta_{\text{max}}$ will be small near threshold and substantial far above threshold, as we have shown. This is the essential ingredient for a sublinear power output.

Our assumption of uniform small-signal gain $g_0$ along the length of the laser is a common one. In electrically injected lasers, this assumption relies on a uniform current injection. This is not necessarily the case in highly-efficient lasers, if the electrical resistance of the laser decreases significantly with an increasing rate of stimulated emission. In such a laser, the current density would concentrate in the regions where the gain is most highly saturated. This effectively delivers electrons to where they are needed most, and would oppose the limiting influence of $\beta_{\text{max}}$. One could also intentionally deliver more current to the region of highest intensity using multiple electrical contacts [64], or taper the waveguide width to dilute the optical intensity and prevent it from nearing $\beta_{\text{max}}$ [64]. Both strategies would mitigate the sublinear power output.
Chapter 5

The quantum cascade laser as a self-pumped parametric oscillator

5.1 Abstract

We report the observation of a clear single-mode instability threshold in continuous-wave Fabry-Perot quantum cascade lasers (QCLs). The instability is characterized by the appearance of sidebands separated by tens of free spectral ranges (FSR) from the first lasing mode, at a pump current not much higher than the lasing threshold. As the current is increased, higher-order sidebands appear that preserve the initial spacing, and the spectra are suggestive of harmonically phase-locked waveforms. We present a theory of the instability that applies to all homogeneously-broadened standing-wave lasers. The low instability threshold and the large sideband spacing can be explained by the combination of an unclamped, incoherent Lorentzian gain due to the population grating, and a coherent parametric gain caused by temporal
population pulsations that changes the spectral gain line shape. The parametric term suppresses the gain of sidebands whose separation is much smaller than the reciprocal gain recovery time, while enhancing the gain of more distant sidebands. The large gain recovery frequency of the QCL compared to the FSR is essential to observe this parametric effect, which is responsible for the multiple-FSR sideband separation. We predict that by tuning the strength of the incoherent gain contribution, for example by engineering the modal overlap factors and the carrier diffusion, both amplitude-modulated (AM) or frequency-modulated emission can be achieved from QCLs. We provide initial evidence of an AM waveform emitted by a QCL with highly asymmetric facet reflectivities, thereby opening a promising route to ultrashort pulse generation in the mid-infrared. Together, the experiments and theory clarify a deep connection between parametric oscillation in optically pumped microresonators and the single-mode instability of lasers, tying together literature from the last 60 years.

5.2 Introduction

In the last decade, significant efforts have spurred the understanding of high-Q optically-pumped microresonators. A monochromatic external pump beam is coupled to a mode of the microresonator, and at sufficient pump power the third-order $\chi^{(3)}$ Kerr nonlinearity, responsible for the intensity-dependent refractive index, couples the pumped mode to fluctuations at other frequencies, which leads to interesting physics. Starting from an initial demonstration of third-order optical parametric oscillation (OPO) [70,71], in which the pump beam provides sufficient parametric gain to allow a few pairs of sidebands to oscillate, this technique has been extended to generate
wide-spanning frequency combs [72,73], and most recently temporal solitons [74,75].
The many degrees of freedom one can manipulate in these systems, such as the group
velocity dispersion (GVD), the free spectral range of the resonator, the detuning of
the pump frequency relative to the cold cavity mode that it pumps, and the pump
to power, among others, have provided a rich nonlinear optical playground to observe
diverse physical phenomena.

A laser, much like an OPO, is an optical resonator in which circulating monochro-
matic light reaches high intensity, the difference being that the light is internally
generated rather than externally injected. Furthermore, the very gain medium that
allows for lasing, simultaneously provides a third-order nonlinearity, the population
pulsation (PP) nonlinearity [76]. The PP nonlinearity is an intrinsic property of any
two-level system that interacts with near-resonant amplitude-modulated (AM) light:
the radiative transition rate between the states, and therefore the population of each
state, is temporally modulated by the AM light, resulting in so-called population pul-
sations that act back on the light field in a nonlinear way. The laser therefore contains
the two ingredients, high-intensity light and a non-linearity, necessary for parametric
oscillation. Indeed, in the late 1960s the importance of PPs in determining the above-
threshold spectral evolution of a homogeneously broadened, traveling-wave laser was
realized. At the laser threshold, one mode—which we call the primary mode–begins
to lase and as the current is increased the population inversion remains clamped to
its threshold value. It was first thought that this clamping should prevent any other
mode from reaching the oscillation threshold. This reasoning, however, neglects the
fact that when a photon of a different frequency is spontaneously emitted in the
presence of the primary lasing field, a beat note—i.e., an intensity modulation at the
difference frequency of the two fields—is created. The beat note creates a PP that
provides a parametric contribution to the gain of the spontaneously emitted photon.
At a sufficiently high pumping level known as the instability threshold, this paramet-
ric gain can—despite the fact that the population inversion is clamped to its threshold
value—allow two sidebands to overcome the loss. The separation of these sidebands
from the primary mode is related to the Rabi frequency induced by the primary mode.
This effect is responsible for both the Haken-Risken-Schmid-Weidlich (HRSW) insta-
bility [77, 78] and the Risken-Nummedal-Graham-Haken (RNGH) instability [79, 80].
Many years later, insightful work properly identified the fundamental role of PPs in
the single-mode laser instabilities [81–87] and also chaos [88]. (We note that PPs are
important not only for inverted media. Historically, their effects were first appreci-
ated in microwave spectroscopy pump-probe experiments by Autler and Townes [89]
in 1955, and soon came to be known as the ac Stark effect. Through the late 1960s and
1970s, significant work on sideband amplification [90], resonance fluorescence [91,92],
and the Mollow scattering triplet [93–97] culminated in the “dressed” description of
atoms in strong fields [98]. In the 1980s, the PP nonlinearity was cast in the language
of nonlinear optics and applications such as four-wave-mixing (FWM) [99], phase
conjugation [100], and optical bistability [83] were explored.)

Both the HRSW and RNGH single-mode instabilities apply to homogeneously-
broadened traveling-wave lasers, and predict the appearance of sidebands on the
primary lasing mode, as shown in Fig. 5.1(a). We remark that in general, the temporal
behavior of an electric field that contains three equally spaced frequencies can be
Figure 5.1: (a) The emission spectrum at the instability threshold comprises a primary mode and two weak sidebands. (b) The temporal behavior of the field $E(t)$ depends on the relative phases of the three modes $\phi_-$, $\phi_0$, and $\phi_+$, and shown are the AM and FM configurations. (c) The AM and FM fields can be understood in terms of the constructive and destructive addition of two beat note phasors, where each phasor represents a contribution to the intensity modulation at the difference frequency $\delta \omega$ resulting from the superposition of each sideband with the primary mode. (d) In a standing-wave cavity, the intensity of each mode varies with position, and the spatial modes corresponding to different frequencies do not perfectly overlap.

More amplitudemodulated or frequency-modulated (FM), depending on the spectral phase, as shown in Fig. 5.1(b). One can think of the intensity modulation (in other words, the beat note) of the AM and FM fields as resulting from the sum of two phasors rotating at frequency $\delta \omega$, each of which is created by the beat between a sideband and the primary mode. As shown in Fig. 5.1(c), the two phasors either constructively interfere to create a large intensity modulation (AM) or destructively interfere to eliminate the intensity modulation (FM). In both the HRSW and RNGH
instabilities, the three-wave field is by necessity AM; a constant-intensity FM field would not create the PP and the resulting parametric gain that is required by the sidebands to reach the lasing threshold. In the HRSW case, which applies to low quality-factor cavities for which the photon lifetime is shorter than the atomic decay time, the sideband separation is smaller than the mode spacing, or free spectral range (FSR), of the cavity. All three lasing frequencies fall within a single cold cavity resonance, which is made possible by a region of anomalous dispersion created by the PP [81]. In the RNGH instability, which applies to higher quality-factor cavities, the sidebands must coincide with cold cavity modes in order to satisfy the roundtrip phase condition, resulting in a separation that is an integer multiple of the FSR.

An important corollary of this requirement is that to observe the effect of the PPs, the FSR must be smaller than the gain recovery frequency (i.e., inverse of the gain recovery time $T_1$). Why? The gain recovery time determines the fastest time scale at which the population inversion can respond to an intensity modulation, therefore the amplitude of the PP is only significant for sidebands detuned by an amount close to or smaller than $1/T_1$. If the FSR is greater than $1/T_1$, it is not possible to simultaneously satisfy the roundtrip gain and phase conditions for a sideband to become unstable. Ideally, the FSR should be significantly smaller than $1/T_1$ so that the FP modes densely populate the parametric gain lobe, increasing the probability of satisfying the instability condition. Provided this condition is met, the RNGH instability predicts that a traveling-wave laser with rapid dephasing must be pumped nine times above threshold before the instability appears. Experimental observations of a rhodamine dye ring laser [85] showed signatures of an RNGH-like instability, with
two key differences: the instability threshold was only fractionally higher—not nine times higher—than the lasing threshold, and the sideband creation was accompanied by the disappearance of the primary mode. Efforts to explain the discrepancies between theory and experiment are well-summarized in [86,101], but to our knowledge the discrepancy was never fully resolved.

In this work, we will investigate the single-mode instability in a standing-wave laser, shown schematically in Fig. 5.1(d). The distinguishing feature of the standing-wave laser is that the primary mode induces a population grating (PG) (as long as carrier diffusion is limited), an effect known as spatial hole burning (SHB). The gain of other cavity modes is no longer clamped above threshold, but continues to increase with the pumping. Therefore, the instability threshold can be reached without the need for PP parametric gain. We call this an incoherent instability, and it occurs in media whose gain recovery time is too slow for PPs to occur (FSR > 1/T₁), such as diode lasers. In gain media with a fast recovery time (FSR < 1/T₁), both the incoherent gain and the parametric PP contribution to the gain must be considered. The PP parametrically suppresses the gain of nearby sidebands, because low-frequency sidebands cause the population inversion to oscillate perfectly out of phase with the intensity modulation. On the other hand, the PP enhances the gain of larger-detuning sidebands, as occurs in the RNGH instability. Depending on the relative contributions of the incoherent and coherent gain, we show that the laser will either emit an FM or an AM waveform at the instability threshold, to either minimize or maximize the amplitude of the PPs. If the incoherent gain is large, nearby sidebands are favored and will yield FM emission to minimize the amount of parametric suppression.
Chapter 5: The quantum cascade laser as a self-pumped parametric oscillator

If the incoherent gain is small, larger-detuning sidebands are favored and will yield AM emission to maximize the amount of parametric enhancement. The possibility of both FM and AM emission from a standing-wave laser is a novelty not shared by the traveling-wave laser, which, as we mentioned, can only produce an AM waveform.

The quantum cascade laser (QCL) is precisely the kind of laser for which both the PG and PPs are important. An electron injected into the upper state has only a short picosecond lifetime during which to diffuse before it scatters to the ground state—not enough time to traverse the half-wavelength mid-infrared ($\lambda \sim 3\text{-}12\mu\text{m}$) standing-wave from node to antinode. Therefore, the PG is not washed out. Also, the FSR (typically 8 to 16 GHz) is much less than the gain recovery frequency ($1/T_1 \approx 1\text{ THz}$), so the population inversion has no difficulty following the beat notes in field intensity created when multiple modes lase simultaneously, yielding PPs. We report the discovery that continuous-wave (cw) Fabry-Perot (FP) QCLs reach a well-defined instability threshold, characterized by the appearance of sidebands whose separation from the primary mode can be several multiples of the cavity FSR. This mode skipping is a clear signature of the parametric PP interaction between the primary mode and the sidebands, which strongly suppresses sidebands at separations much smaller than the large gain recovery frequency of the QCL. The behavior is observed in QCLs that emit at wavelength 3.8\,$\mu$m, 4.6\,$\mu$m, and 9.8\,$\mu$m, indicating that it is a universal feature of mid-infrared QCLs, independent of the specific bandstructure of the active region.

The strength of the PG can be tuned by coating the facets to adjust their reflectivities. By comparing the measurements with the theory, we argue that QCLs with uncoated facets emit an FM waveform. A QCL with one high-reflectivity facet and a sufficiently
low reflectivity of the other facet should in principle emit an AM waveform, and we provide preliminary evidence that this is indeed the case, demonstrating a QCL whose sidebands are separated from the primary mode by 46 FSR. While the PG and PP have been known to be important in QCLs, in previous work their effects were treated separately [102]. Instead, we emphasize that one should think of the PG—a spatial modulation of the inversion—and the PP—a temporal modulation of the inversion—as working in tandem to create a phase-locked multimode state at low pump power.

As the current is increased past the instability threshold, higher-order sidebands that preserve the initial spacing appear. This suggests that the FP-QCL can emit a harmonically phase-locked waveform without the need for any external modulation or additional nonlinear elements. Why have such spectra not been observed before, except in a few cases [103, 104]? We have found the harmonic states to be extremely sensitive to optical feedback. Simply placing a collimating lens between the QCL and the spectrometer—even a poorly aligned, tilted lens with a focal length of a few cm—makes it difficult to observe the harmonic state, and instead yields the more familiar QCL spectrum in which all adjacent FP modes lase. It is also important to slowly increase the current, which allows for a smooth transition from the single-mode to the harmonic regime. We argue that the harmonic state is an intrinsic regime of all QCLs. The fact that it has only been observed 15 years after the invention of the cw QCL is a testament to the destabilizing influence of optical feedback [105].

In the last few years, comb generation in a QCL on adjacent FP modes has been demonstrated [106–108], and the importance of parametric mode coupling is known [109, 110]. (These devices all had multi-stage inhomogeneously broadened
active regions, which distinguishes them from the devices in our work.) Because these combs have so far always comprised adjacent cavity modes, consideration has only been given to the case where the fundamental frequency of the PPs equals the FSR. This low PP frequency strongly favors the emission of an FM waveform. The remarkable degree of freedom to skip modes, never before considered, means that the temporal periodicity of the PPs is no longer pinned to the cavity roundtrip time (typically 60 to 120 ps), but is shortened by a factor equal to the number of modes skipped, which reaches 46 in one of our QCLs. This reduction of the period down to the order of the gain recovery time is the crucial feature that allows for the possibility of AM emission.

Finally, we emphasize the deep connection between the single-mode laser instability and mode proliferation in optically pumped microresonators. Both are cases of parametric oscillation that are initiated by a nonlinearity, either PP or Kerr, transferring energy from a pump beam to two sidebands. For a passive microresonator the pump beam must be injected, while in the laser the pump beam is internally generated. This analogy, which we only begin to uncover here, can help guide future work toward understanding the rich emission spectra of QCLs. We hope that the advancement of the QCL can parallel the rapid progress seen in microresonators in the last decade, leading to a compact source of mid-infrared frequency combs for dual-comb spectroscopy of trace gases and short pulse generation [111].

In Sec. 5.3 we present the experimental results, which helps to motivate the theory presented in Sec. 5.4. In Sec. 5.5 we compare the theory with the measurements, and finally conclude in Sec. 5.6.
5.3 Experiment

All four devices used in this study are cw, buried heterostructure, FP-QCLs. Our device naming convention identifies the provider of the device (LL: MIT Lincoln Laboratory, TL: Thorlabs, DS: Daylight Solutions) followed by the emission wavelength in microns. The active region of device LL-9.8 is a double phonon resonance design using lattice-matched Ga$_{0.47}$In$_{0.53}$As/Al$_{0.48}$In$_{0.52}$As, grown by metalorganic chemical vapor deposition, with the well-known layer structure of [112], albeit with a nominal doping of $n = 2.5 \times 10^{18}$ cm$^{-3}$, for which extensive bandstructure calculations have been done [113]. The device length is 3 mm and width is 8 µm. Devices TL-4.6, TL-4.6:HR/AR, and DS-3.8 were grown using strained Ga$_x$In$_{1-x}$As/Al$_y$In$_{1-y}$As and are described in [114], although the layer sequence is not given. The length is 6 mm and width is 5 µm for these three devices. Both facets are left uncoated for LL-9.8, TL-4.6, and DS-3.8. The only coated device is TL-4.6:HR/AR, which has a high-reflectivity (HR) coating on the back facet ($R \approx 1$) and an antireflection (AR) coating on the front facet ($R \approx 0.01$), but is otherwise nominally identical to TL-4.6. Far-field measurements indicate that all devices exhibit single lateral-mode emission over the full range of applied current. It is worth mentioning that the short-wave QCLs, DS-3.8 and TL-4.6, have positive GVD and the long-wave device LL-9.8 has negative GVD. We expect this because their wavelengths lie on opposite sides of the zero-GVD point of InP, but we have also confirmed this using the measurement method of [115]. Some relevant parameters for each device are given in Table 5.1: the effective refractive index $n_{eff}$ is determined from the FP-mode spacing of the measured spectra; the dipole moment $d$ and the upper state lifetime $T_{up}$ are calculated from the
Figure 5.2: (a) Total power output of each QCL (from both facets) vs. current, color-coded to indicate the range over which the laser operates in a single-mode, harmonic state, or dense state. (b) The intracavity power normalized to the saturation intensity (calculated from the measured output power and the best estimates for $\kappa$, $T_1$, $T_2$, and the facet reflectivities) is plotted vs. $J/J_{th}$.

Our goal was to precisely examine the spectral evolution of the QCL with increasing current, from the single-mode to the multimode regime. Specifically, we wanted to answer the question: at what pumping level does a second mode start to lase, and what is the relationship between the second frequency and the first? To
Table 5.1: Summary of relevant parameters of the devices used in this study. 

answer this question, we would begin each measurement with the laser driven at a 
current beneath the laser threshold. The current was then slowly increased in steps 
of 1 mA, and the spectrum was monitored using a Fourier transform infrared (FTIR) 
spectrometer (Bruker Vertex 80v), with either a nitrogen-cooled InSb detector (for 
DS-3.8 and TL-4.6) or HgCdTe detector (for LL-9.8). The current was supplied by 
a low-noise driver (Wavelength Electronics QCL1500 or QCL2000), and the temper-
ature of the copper block beneath the QCL was stabilized to 15°C. The slow rate of 
increase of the current was necessary to precisely identify the instability threshold, 
and also to prevent rapid temperature variations. To completely eliminate the pos-
sibility of optical feedback due to reflections from optical elements outside the laser 
cavity, the QCL was placed about 40 cm from the entrance window to the FTIR and 
its output was not collimated with a lens, but simply allowed to diverge. The high 
power of the devices and the sensitivity of the detectors was sufficient to measure 
spectra despite the small fraction of collected optical power.

Spectra measured in this manner are shown in Fig. 5.3(a-c) for the three un-
coated devices. Each spectrum is normalized to its own maximum and plotted on
Figure 5.3: Spectra of the three uncoated QCLs (a) LL-9.8, (b) TL-4.6 and (c) DS-3.8 as the current is incremented, starting from below threshold, and also (d), (e), (f) as the current is decremented, starting at the maximum current that was reached on the upward current ramp.
a logarithmic scale covering 40 dB of intensity variation. All three lasers undergo a very similar spectral evolution. Above threshold, the laser remains single-mode for a substantial range of current until a clear instability threshold is reached, at which a 1 mA increase in current results in the appearance of new lasing modes. The new frequencies appear as symmetric sidebands on the primary lasing frequency, with a separation that is many integer multiples of the FSR. The sideband spacing $\delta \omega_{sb}$ and pumping $J_{sb}$ at the sideband instability threshold are given in Table 5.1 for each device. Taking LL-9.8 as a first example, at $J_{sb}/J_{th} = 1.14$ a pair of equal-amplitude sidebands separated by 7 FSR from the primary mode suddenly rise out of the noise floor to an intensity 20 dB weaker than the primary mode. As the current increases further, higher-order sidebands appear that preserve the initial spacing, eventually yielding a spectrum at $J/J_{th} = 1.39$ of 11 modes, each separated by 7 FSR from its nearest neighbors. We refer to a spectrum of modes separated by multiple FSR as a harmonic state. Above $J/J_{th} = 1.39$, interleaving modes incommensurate with the harmonic spacing begin to appear. At $J/J_{th} = 1.47$, there is another sudden transition at which all adjacent FP modes are populated; we refer to this as a “dense” state, and it persists as the current is increased to roll-over. For device TL-4.6, sidebands with a separation of 26 FSR from the primary mode appear at $J_{sb}/J_{th} = 1.17$, and the transition to the dense state occurs at $J/J_{th} = 1.30$. For device DS-3.8, the sideband separation is 20 FSR at $J_{sb}/J_{th} = 1.12$. As the current is increased, the sideband spacing displays a sudden jump from 20 FSR to 25 FSR. At $J/J_{th} = 1.32$ the laser jumps to a dense state for somewhere between a few seconds and a minute before returning to a “noisy” harmonic state: one with prominent harmonic peaks but many
incommensurate modes populated as well. At $J/J_{th} = 1.38$ the dense state appears again, and this time persists for all higher currents.

After the laser enters the dense state, we decrease the current slowly and study the spectral evolution. There is a remarkable hysteresis, as shown in Fig. 5.3(d-f). For LL-9.8, the dense state persists all the way until $J/J_{th} = 1.01$, when the single-mode finally reappears. A similar hysteresis occurs in TL-4.6 and DS-3.8. For DS-3.8, a noisy harmonic state can appear, and the laser can jump from a noisy harmonic state back to a dense state as the current is decreased further. A general observation for all three devices is that the clean harmonic state cannot be recovered once the laser has entered the dense state.

Lastly, we present in Fig. 5.4(a) the spectral evolution of TL-4.6-HR/AR as the current is incremented. The behavior of this device is different from the uncoated devices in two significant ways: 1) the sidebands appear with a separation of 46 FSR,
Chapter 5: The quantum cascade laser as a self-pumped parametric oscillator

much larger than any spacing seen previously, at \( J_{sb}/J_{th} = 1.22 \), and 2) the harmonic regime persists over a much larger range of output power than it does in the uncoated devices, as seen by the color-coding in Fig. 5.2. (A second device, nominally identical to TL-4.6:HR/AR, developed sidebands with a spacing of 48 FSR at \( J_{sb}/J_{th} = 1.18 \).) As the current is decreased, as shown in Fig. 5.4(b), the dense state persists until a noisy harmonic state appears at \( 1.29J_{th} \). Interestingly, in this device—unlike the uncoated ones—the clean harmonic state with one pair of sidebands reappears at 885 mA, which is quite close to the instability threshold of 880 mA found when the current is ramped upwards.

When the spectral evolution measurement is repeated many times for one device, starting from below threshold and incrementing the current, we find that the instability threshold \( J_{sb} \) and sideband spacing \( \delta \omega_{sb} \) are always the same. As the current is increased past \( J_{sb} \), there can be slight variations from one experiment to another. For example, the jump from 20 to 25 FSR in TL-4.6 does not always occur at the exact same current, but predictably within a range of about 20 mA. The same is true of the transition to the dense state.

The IV curve of device DS-3.8 shown in Fig. 5.5 demonstrates a hysteresis that is correlated with the spectral hysteresis. When starting below threshold and increasing the current (red), the voltage of the laser decreases (negative differential resistance) when the noisy harmonic state transitions to the dense state at 523 mA. (In the spectra of DS-3.8 shown in Fig. 5.3(c) of the manuscript, this transition occurs at 502 mA. The exact current at which the transition occurs is not identical from one experiment to another, but occurs predictably within a range of about 20 mA.) Once
Figure 5.5: The IV curve of DS-3.8 exhibits a hysteresis as the current is increased (red) and decreased (blue). The hysteresis is correlated with the transition from the noisy harmonic state to the dense state.

When the laser reaches the dense state and the current is subsequently decreased (blue), the laser persists in the dense state at currents below 523 mA. This is responsible for the hysteresis loop. For the same current, the voltage is 4.8 mV smaller when in the dense state than in the noisy harmonic state. At 494 mA, the dense state transitions to the noisy harmonic state, and the two voltage curves overlap again. The lower voltage of the dense state indicates a larger radiative photocurrent, which implies that the output power of the laser is slightly larger in the dense state than in the harmonic state, for the same current pumping. (While we could have measured the output power to demonstrate this, the IV measurement is more sensitive.) Thus, the dense state is more efficient at extracting photons from upper state electrons and is, in one sense, more stable.
We would like to know the temporal behavior of the emitted field in the harmonic state, specifically whether it has more AM or FM character. Unfortunately this cannot be determined by measuring the strength of the beat note of the laser output, because the smallest observed beat frequency is greater than 100 GHz, larger than the electrical bandwidth of any mid-infrared photodetector. We therefore look to theory for insight on the matter, and plan on second-order autocorrelation experiments in future work.

5.4 Theory

The instability threshold is characterized by the appearance of symmetric sidebands on the primary lasing mode. Our goal is to theoretically explain the frequency separation of the sidebands and the pump power at which they first appear. We begin with the general framework: the Maxwell-Bloch equations for a two-level system and the spatial mode expansion of a laser cavity. Then, we first address the single-mode solution of the laser to determine how the primary mode and the population grating evolve with increasing pumping. Secondly, we must understand how the two-level system responds to a weak field at a frequency different from that of the primary mode—the population pulsation. Finally, we will combine these two ingredients, the PG and the PP, to explain the instability threshold. We find that the PG provides an unclamped Lorentzian contribution to the gain of the sidebands, which is responsible for the low instability threshold. The PP reshapes the gain, suppressing nearby sidebands and enhancing more distant ones, and is responsible for the observed multiple-FSR sideband separation. Interestingly, we find that depending on
the relative contributions of the PG and the PP to the gain, the laser can emit either an FM or AM waveform at the instability threshold.

5.4.1 General Framework

We model the lasing transition as a two-level system, or a quantum dipole, subject to the electric field

\[ E(t) = \mathcal{E}(t)e^{i\omega_0 t} + \text{c.c.} \quad (5.1) \]

The response is characterized by the population inversion \( w \) (positive when inverted) and the off-diagonal element of the density matrix \( \sigma \), which in turn obey the Bloch equations (in the rotating wave approximation) [117],

\[ \dot{\sigma} = \left( i\Delta - \frac{1}{T_2} \right) \sigma + \frac{i\kappa}{2} w \mathcal{E} \quad (5.2) \]

\[ \dot{w} = i\kappa (\mathcal{E}^* \sigma - \mathcal{E} \sigma^*) - \frac{w - w_{eq}}{T_1} \quad (5.3) \]

where \( \Delta = \omega_{ba} - \omega_0 \) is the detuning between the field and the resonant frequency \( \omega_{ba} \) of the two-level system, \( T_1 \) is the gain recovery time, \( T_2 \) is the dephasing time, \( \kappa \equiv 2d/\hbar \) is the coupling constant where \( d \) is the dipole matrix element (assumed to be real) and \( \hbar \) is Planck's constant, and \( w_{eq} \) is the "equilibrium" population inversion that the system would reach in the absence of photons, determined by the pumping. (Note that we have defined \( T_1 \) to be the gain recovery time, which in QCLs is distinct from the upper state lifetime \( T_{up} \) due to the nature of electron transport in the active region. See Sec. 5.8 for a discussion of this subtlety.) We write the macroscopic polarization \( P \) (dipole moment per volume) as

\[ P(t) = \mathcal{P} e^{i\omega_0 t} + \text{c.c.}, \quad (5.4) \]
where $\mathcal{P} = N d \sigma$, and $N$ is the volume density of dipoles.

A characteristic of the two-level medium that will appear often is the “Beer loss rate”

$$\tilde{\alpha} = \frac{N d^2 T_2 \omega_0 c \sqrt{\mu/\epsilon}}{\hbar},$$

which is related to the more familiar Beer absorption coefficient $\alpha$ (with units of inverse length) that appears in Beer’s law of absorption by $\tilde{\alpha} = \alpha c$. We adopt the convention of [45] and assume our dipoles to be embedded in a host medium of permittivity $\epsilon$ and permeability $\mu$. The speed of light $c = 1/\sqrt{\epsilon \mu}$ also denotes the value in the background host medium.

In the standing-wave cavity, the field envelopes vary in space in addition to time. We follow a common approach and decouple the spatial and temporal dependence, writing the field as

$$E(z, t) = \sum_{m=-0,+} \Upsilon_m(z) \mathcal{E}_m(t) e^{i \omega_m t} + c.c.,$$

where $\omega_0$ is the primary mode frequency and the sideband frequencies are $\omega \pm = \omega_0 \pm \delta \omega$. These three frequencies are cold-cavity resonant frequencies, and are equidistant from one another because we have assumed zero GVD. We will henceforth assume that the primary mode $\omega_0$ lases at the resonant frequency of the two-level system, so $\Delta = 0$. This is a reasonable approximation if the FSR is much smaller than the gain bandwidth. These two assumptions, $\text{GVD} = 0$ and $\Delta = 0$, simplify later mathematical formulas considerably and allow for easier understanding of the essential physics. The full theory without these assumptions is included in Sec. 6.3. The spatial modes $\Upsilon_m(z)$ are determined by the cavity geometry, and do not vary in time. We assume the laser cavity to have mirrors with unity reflectivity, so that the spatial
modes are given by

\[ \Upsilon_m(z) = \sqrt{2} \cos(k_m z) \]  

(5.7)

where \( k_m \) is an integer multiple of \( \pi / L \) and \( L \) is the length of the cavity. The mirror loss \( \ln(1/\sqrt{T_1 T_2})/L \) is included in the total optical losses of the cavity, \( \tilde{\ell} \). The assumption of perfect reflectivity simplifies the problem in two important ways: the spatial functions \( \Upsilon_m(z) \) are orthogonal, and they do not change shape as the pumping increases. This assumption turns out to be quite good even for semiconductor lasers with facet reflectivities around 0.25. The approximation breaks down for our HR/AR coated QCL, and here we will only give a qualitative description of what happens and save the considerably more complicated theory for a future publication.

### 5.4.2 Population Grating

The threshold inversion is given by the ratio of the optical loss rate to the Beer loss rate, \( w_{th} = \tilde{\ell}/\tilde{\alpha} \). We define the pumping parameter \( p \equiv w_{eq}/w_{th} \). When \( p = 1 \), the primary mode begins to lase at the frequency \( \omega_0 = \omega_{ba} \). As the pumping \( p \) is increased, the field and inversion can be solved for by the method of [102], which is detailed in Sec.6.2. We account for the population grating, but not the coherence grating which has been incorporated in recent work [118]. The primary mode \( \tilde{E}_0 \) grows according to

\[ |\tilde{E}_0|^2 = \frac{p - 1}{1 + \gamma_D/2}, \]  

(5.8)

where we have defined the dimensionless primary mode amplitude \( \tilde{E}_0 \) by normalizing by the saturation amplitude, \( \tilde{E}_0 \equiv \kappa \sqrt{T_1 T_2} E_0 \). The diffusion parameter \( \gamma_D \) is given by \( \gamma_D = (1 + 4k_0^2 D T_{up})^{-1} \), where \( D \) is the diffusivity of the excited-state electrons

105
and $T_{up}$ is the upper-state lifetime. The parameter $\gamma_D$ ranges from 0 (for infinite mobility) to 1 (for zero mobility). The population inversion in the presence of the primary mode, $w_0(z)$, varies with $p$ as

$$w_0(z) = w_{th} \left[ 1 + \frac{\gamma_D}{2} \frac{p - 1}{1 + \gamma_D/2} - \gamma_D \frac{p - 1}{1 + \gamma_D/2} \cos(2k_0z) \right].$$  \hspace{1cm} (5.9)$$

Equations 5.8 and 5.9 are valid to first order in the primary mode intensity $|\tilde{E}_0|^2$, or equivalently, $p - 1 \ll 1$. Note that for zero diffusion ($\gamma_D = 1$), the slope efficiency of the laser is two thirds that of the infinite diffusion ($\gamma_D = 0$) case. This is because for infinite diffusion, the inversion is uniformly pinned to $w_{th}$ above threshold. For finite diffusion, as the pumping increases the population grating grows in amplitude. At the same time, the average value of the inversion increases, indicating that the inversion is not being converted into photons as efficiently as it could be if the electrons could diffuse from the field nodes to the antinodes. (In principle, one can extract $\gamma_D$ from measurements of the slope of $|\tilde{E}_0|^2$ vs. $p$, which should be between zero and one. Figure 5.2(b) shows $|\tilde{E}_0|^2$ vs. $J/J_{th}$. All curves have a slope greater than one, which suggests that $J/J_{th}$ is an underestimate of $p$. See Sec. 5.7 for how $|\tilde{E}_0|^2$ is determined from the measurements, and how the transparency current can cause $J/J_{th}$ to underestimate $p$. Therefore, more characterization is needed to extract $\gamma_D$ from the measurements.)

### 5.4.3 Population Pulsation

To understand the population pulsation, we can ignore the spatial dependence of the intracavity field and consider only a single two-level system subject to an applied
field

$$E(t) = \sum_{m=-,0,+} \mathcal{E}_m(t)e^{i\omega_m t} + \text{c.c.} \quad (5.10)$$

Since we are interested in calculating the stability of the sidebands, the amplitudes $\mathcal{E}_\pm$ should be thought of as infinitesimal perturbations; as such, our entire treatment retains only terms to first order in the sideband amplitudes $\mathcal{E}_\pm$. We write the total polarization as

$$P(t) = \sum_{m=-,0,+} \mathcal{P}_m(t)e^{i\omega_m t} + \text{c.c.} \quad (5.11)$$

The polarization at the sidebands can be calculated using Eqs. 5.2 and 5.3 [119], which gives

$$\mathcal{P}_+ = \frac{i\epsilon}{\omega_{ba}} \tilde{\alpha} w_0 \left[ \frac{\mathcal{E}_+}{1 + i\delta\omega T_2} + \Lambda \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0^* \mathcal{E}_+ + \Lambda \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0 \mathcal{E}_-^* \right] \quad (5.12)$$

$$\mathcal{P}_- = \frac{i\epsilon}{\omega_{ba}} \tilde{\alpha} w_0 \left[ \frac{\mathcal{E}_-}{1 - i\delta\omega T_2} + \Lambda^* \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0^* \mathcal{E}_- + \Lambda^* \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0 \mathcal{E}_+^* \right], \quad (5.13)$$

where

$$\Lambda = \frac{-\left(1 + i\delta\omega T_2/2\right)}{\left((1 + i\delta\omega T_1)(1 + i\delta\omega T_2)^2 + (1 + i\delta\omega T_2)|\tilde{\mathcal{E}}_0|^2\right)} \quad (5.14)$$

and $w_0$ is the saturated population inversion $w_0 = w_{eq}/(1 + |\tilde{\mathcal{E}}_0|^2)$. The polarization at each sideband is neatly divided into three contributions. Taking $\mathcal{P}_+$ as an example, the first term in Eq. 5.12 is the Lorentzian contribution that the sideband generates due to the linear susceptibility of the dipole. The second and third terms are nonlinear contributions due to the PP at frequency $\delta\omega$: a self-mixing term of the sideband with the primary mode, and a cross-mixing term of the other sideband with the primary mode. The frequency-dependent portion of the nonlinear susceptibility is $\Lambda$, which is a dimensionless function of the sideband detuning, the time constants of the two-level system, and the primary mode intensity.
From the field and the induced polarization, we can calculate the total power density generated, $\langle -E\hat{P} \rangle$. The quantity that most interests us is the gain $\bar{g}$ (with dimension of frequency) seen by each sideband, defined as the power generated at the sideband’s frequency, divided by the energy density of the exciting sideband field.

To develop a feel for the parametrically generated polarization and the resulting gain, we consider two instructive cases. In both cases we take the sidebands to have equal magnitudes, $|\mathcal{E}_+| = |\mathcal{E}_-|$, but choose the phases of the sidebands to give rise to an AM waveform in one case and a constant-intensity FM waveform in the other case, as shown in Fig. 5.1(b). The gain of each sideband is found to be

$$
\bar{g} = \bar{\alpha}w_0 \left[ \frac{1}{1 + (\delta\omega T_2)^2} + \text{Real}(\Lambda) |\tilde{\mathcal{E}}_0|^2 \right] \begin{cases} 2 & \text{AM} \\ 0 & \text{FM} \end{cases}.
$$

(5.15)

The first term is the Lorentzian contribution to the gain, and the second term is the parametric gain due to the PP. The factors of two and zero come from the constructive or destructive addition, respectively, of the self-mixing and cross-mixing terms to the nonlinear polarization. Equivalently, one can say that the constant-intensity FM field does not create a PP, and accordingly experiences no parametric gain. The parametric gain of the AM field is proportional to $\text{Real}(\Lambda)$ and to the primary mode intensity $|\tilde{\mathcal{E}}_0|^2$. (We note that one can quickly derive the original RNGH instability for a traveling-wave laser from Eq. 5.15, which is done in Sec. 6.4.1.) By expanding $\Lambda$ in Eq. 5.14 in powers of $|\tilde{\mathcal{E}}_0|^2$, it becomes clear that the PP interaction can be expressed in the perturbative expansion of traditional nonlinear optics as a third, fifth, seventh, etc. order nonlinearity. We will later calculate the instability threshold in the limit of small primary mode intensity, and are therefore interested in the lowest-
order nonlinearity. We obtain $\chi^{(3)}$, the dimensionless frequency-dependent portion of the third-order PP nonlinear susceptibility, by evaluating $\Lambda$ at $|\tilde{E}_0|^2 = 0$,

$$\chi^{(3)} = \frac{-(1 + i\delta\omega T_2/2)}{(1 + i\delta\omega T_1)(1 + i\delta\omega T_2)^2}. \quad (5.16)$$

To better elucidate the nature of the PPs, the magnitude, phase, and real part of $\chi^{(3)}$ are plotted in Fig. 5.6 as a function of the sideband detuning, for three different values of $T_1/T_2$. At low frequencies $\delta\omega$, the population inversion has no difficulty following the modulation of the field, which has two consequences: the amplitude of the PPs is large, and the PP is $\pi$ out of phase with the intensity modulation of the exciting field. This can be understood simply in terms of rate equations: when the field is stronger, the stimulated emission rate is larger, and the population inversion is therefore smaller. This scenario—higher inversion when the intensity is lower and lower inversion when the intensity is higher—is less efficient at extracting power from the two-level system relative to the case of monochromatic or FM excitation; mathematically, this is described by a parametric gain (determined by the real part of $\chi^{(3)}$) that is negative. We refer to this effect as parametric suppression: a low-frequency PP reduces the gain of each sideband. As $\delta\omega$ increases, the inversion can no longer as easily follow the intensity modulation, so the amplitude of the PPs decreases and the phase of $\chi^{(3)}$ decreases from $\pi$. For large enough $\delta\omega$, the phase of $\chi^{(3)}$ decreases below $\pi/2$, at which point $\text{Real}[\chi^{(3)}]$ becomes positive. We refer to this effect as parametric enhancement: a high-frequency PP increases the gain seen by each sideband. The crossing frequency $\delta\omega_{\text{cr}}$ which separates the low-frequency suppression regime and
Parametric suppression

Parametric enhancement

Figure 5.6: The magnitude, phase, and real part of $\chi^{(3)}$ are plotted vs. $\delta \omega T_2$ for three different ratios $T_1/T_2 = 40, 10, 5$. The parametric gain seen by the sidebands is determined by $\text{Real}[\chi^{(3)}]$. Low-frequency PPs lead to a parametric suppression of the gain. For the gain to be parametrically enhanced, $\delta \omega$ must be large enough that the inversion can no longer follow the intensity in anti-phase; in other words, the phase of $\chi^{(3)}$ must be between $-\pi/2$ and $\pi/2$. 
high-frequency enhancement regime is given by

$$\delta \omega_{cr} T_2 \approx \sqrt{\frac{2/3}{T_1/T_2}},$$

(5.17)

where we have made the approximation $T_1/T_2 \gg 1$, valid for QCLs. The regions of parametric suppression and enhancement are highlighted in the plots of the phase and real part of $\chi^{(3)}$ in Fig. 5.6. Finally, at very large $\delta \omega$ the parametric gain approaches zero (from above), because the beat note becomes too short for the inversion to follow and the amplitude of the PP approaches zero.

It is worth pointing out that in the weak-field limit $|\tilde{E}_0|^2 \ll 1$ that we are interested in, $\delta \omega_{cr}$ has no relation to the Rabi frequency $\Omega_R$ induced by the primary mode,

$$\Omega_R T_2 = \frac{|\tilde{E}_0|}{\sqrt{T_1/T_2}}.$$  

(5.18)

The Rabi frequency of course varies with the primary mode amplitude, while $\delta \omega_{cr}$ is independent of $\tilde{E}_0$ in the weak-field limit. By comparing the factors $\sqrt{2/3}$ and $|\tilde{E}_0|$ in the numerators of Eqs. 5.17 and 5.18, it’s clear that in the limit $|\tilde{E}_0|^2 \ll 1$, $\delta \omega_{cr}$ will always be greater than the Rabi frequency. Thus, the reason for the parametric enhancement when $\delta \omega > \delta \omega_{cr}$ should simply be ascribed to the fact that at high PP frequency, the phase lag between the population inversion and the field intensity becomes appropriate for gain rather than absorption.

5.4.4 Instability Threshold

Now we can ask the question: what happens to the single-mode laser solution when it is perturbed by a weak sideband field? The source of the perturbation could be
spontaneous emission, or even spontaneous parametric downconversion of two primary
type photons into two sideband photons [120]. Our goal is to calculate the gain of the
sideband modes averaged over the length of the cavity. The instability threshold is
reached when the sideband gain equals the loss. Although our instability analysis will
not tell us about the steady-state reached by the sidebands, one reasonable possibility
is that the sidebands begin to lase, as seen in the experimental spectra.

To determine the sideband gain, we start with the polarization in Eqs. 5.12-5.13
and account for the position-dependence by replacing \( \mathcal{E}_m(t) \) with \( \mathcal{E}_m(t) \mathcal{Y}_m(z) \), and \( w_0 \)
with \( w_0(z) \) from Eq. 5.9. In keeping with our approximation to order \(|\tilde{\mathcal{E}}_0|^2\), we replace
\( A \) with \( \chi^{(3)} \). The position-dependent polarization is then inserted as the source term
in Maxwell’s wave equation. From here, the calculation follows the same steps as the
instability analysis done for Kerr microresonators [121], and is detailed in Sec.6.4.

After making the slowly varying envelope approximation, and projecting the equation
onto each of the orthonormal spatial modes, one finds a first order differential equation
for each sideband amplitude. Unlike the earlier example where we hand-picked the
phases of the sidebands to study the effect of an AM and FM field, here the AM and
FM sideband configurations emerge organically as the two “natural modes" of the
system of two sideband equations. The natural modes [83] are the configurations of
the three-wave field for which the relative phases of the fields are preserved as time
evolves; in other words, an AM field remains AM, and an FM field remains FM. (In
the general case of nonzero \( \Delta \) and GVD, the natural modes can be a superposition
of AM and FM.) The gain of the AM and FM natural modes is given by

\[
\bar{g} = \frac{1 + \frac{\gamma_D}{2} \frac{p-1}{1+\gamma_D/2} + \text{Real}[\chi^{(3)}] \frac{p-1}{1+\gamma_D/2}}{1 + (\delta \omega T_2)^2} \cdot \begin{cases} 
\Gamma_{\text{self}} + \Gamma_{\text{cross}} = \frac{3}{2} & ; \text{AM} \\
\Gamma_{\text{self}} - \Gamma_{\text{cross}} = \frac{1}{2} & ; \text{FM}
\end{cases}
\]

(5.19)

where the \( \Gamma \)'s are longitudinal spatial overlap factors

\[
\Gamma_{\text{self}} = \frac{1}{L} \int_0^L dz \left| \gamma_0(z) \right|^2 \left| \gamma_\pm(z) \right|^2 = 1
\]

(5.20)

\[
\Gamma_{\text{cross}} = \frac{1}{L} \int_0^L dz \left\{ \gamma_0(z) \gamma^*_\pm(z) + \gamma_\pm(z) \gamma^*_0(z) \right\} = 1/2.
\]

(5.21)

By comparing the standing-wave sideband gain in Eq. 5.19 to the sideband gain of a single two-level system in Eq. 5.15, we see that the cavity introduces two modifications. First, the Lorentzian gain contribution increases with \( p \); this unclamped gain is a direct result of the PG that develops in the presence of non-zero \( \gamma_D \). Secondly, the partial overlap of the sideband spatial modes \( \gamma_+ \) and \( \gamma_- \) results in partial (rather than complete) interference of the self-mixing and cross-mixing contributions to the gain. To understand this, note from Fig. 5.1(a) that although the emitted waveform has equal-amplitude sidebands, within the cavity the plus and minus sidebands have unequal amplitudes at most positions \( z \), as shown by the red and blue modes in Fig. 5.1(d). Therefore, the self and cross-mixing contributions to the sideband polarization at each position \( z \) cannot completely interfere, and the factors of \( 3/2 \) (AM) and \( 1/2 \) (FM) emerge after averaging over the full cavity length, as opposed to the factors of 2 and 0 in Eq. 5.15. Thus, even when the laser emits an FM waveform, there is still a parametric contribution to the gain due to the incomplete destructive interference of the PP within the cavity.

The instability occurs when \( p \) reaches a value such that the sideband gain \( \bar{g} \) in Eq. 5.19 equals the loss \( \bar{\ell} \) for one particular sideband detuning \( \delta \omega \). (As discussed
Figure 5.7: Overview of the three different types of instabilities. (a) The incoherent instability relies only on the unclamped gain due to the PG, and occurs when parametric effects can be neglected. (b) The FM instability occurs when a strong PG exists, and (c) the AM instability occurs for a weak PG. In (b) and (c), the value $T_1/T_2 = 20$ was used. The FP modes (green) are not associated with the ordinate, and simply provide a sense of the mode spacing.
previously, we assume the FSR is small so that a FP mode always exists very close to the unstable value of $\delta \omega$.) As $p$ increases, the incoherent Lorentzian gain increases, but the parametric gain either increases or becomes more negative depending on the sign of $\text{Real}[\chi^{(3)}]$ (which depends on $\delta \omega$). The three parameters $T_1$, $T_2$, and $\gamma_D$ affect the relative importance of the incoherent and coherent gain terms, and depending on the values of these parameters, one of three different classes of instability can occur: the incoherent instability, FM instability, and AM instability. In Fig. 5.7, each type of instability is illustrated by plotting the sideband gain at the instability threshold, which we now explain.

**Incoherent Instability**

The parametric gain can often be ignored. If $T_1$ is large enough, the interesting features of $\text{Real}[\chi^{(3)}]$ all occur for sideband detunings less than 1 FSR, and so the parametric gain will be nearly zero for all values of $\delta \omega$ greater than 1 FSR. This is the case for diode lasers, where PPs are significant up to a few GHz ($T_1 \approx 1 \text{ ns}$), while the FSR is around 100 GHz. Thus, only the incoherent gain term in Eq. 5.19 matters (although it is not a Lorentzian for bandgap lasers). As $p$ increases beyond 1, the sideband gain increases but remains Lorentzian, so the sidebands that reach the instability threshold first will always be the FP modes immediately adjacent to the primary lasing mode [122]. In diode lasers, $\gamma_D$ is small ($\sim 10^{-4}$), so $p$ needs to be large before the second mode can appear.

The value of $\gamma_D = 0.5$ in Fig. 5.7(a) is typical of short-wave QCLs. We see that if coherent effects were negligible in QCLs, we would expect the sidebands to
appear at \( p = 1.001 \), barely above threshold. The much higher instability threshold measured in the experiments, together with the observation that the sidebands do not appear at the nearest-neighbor FP modes of the primary mode, indicates that coherent effects play an essential role in the QCL instability.

**FM Instability**

When the Lorentzian gain increases quickly with \( p \) due to a strong PG, sidebands that fall within the parametric suppression regime, \( \delta \omega < \delta \omega_{cr} \), can reach the instability threshold. This is counterintuitive: why should a sideband lase when the parametric interaction provides negative gain? The answer is that the Lorentzian gain favors sidebands with as small a separation as possible, and if it is large enough it can pull sidebands above threshold in spite of the negative contribution from the parametric gain. In this scenario, FM sidebands have a lower instability threshold than AM sidebands because the parametric contribution to the gain is less negative, since \( 1/2 < 3/2 \) in Eq. 5.19. Such a case is illustrated in Fig. 5.7(b) for \( \gamma_D = 0.5 \) and \( T_1/T_2 = 20 \). At \( p = 1.06 \), FM sidebands reach the instability threshold, while AM sidebands are too strongly suppressed to reach the instability. A key feature of the instability is that the unstable sideband will be several FSR away from the primary mode (provided that the FSR is small), while still satisfying \( \delta \omega < \delta \omega_{cr} \).

**AM Instability**

When the Lorentzian gain increases little with \( p \) due to a weak PG, only sidebands that fall within the parametric enhancement regime, \( \delta \omega > \delta \omega_{cr} \), will be able to reach the instability. In this case, AM will have a lower instability threshold than FM.
because AM receives a larger parametric enhancement (since $3/2 > 1/2$ in Eq. 5.19). Such a case is illustrated in Fig. 5.7(c) for $\gamma_D = 0.04$ and $T_1/T_2 = 20$. At $p = 1.9$, AM sidebands reach the instability threshold while the FM sidebands are not sufficiently enhanced to reach the instability. Strictly speaking, $p=1.9$ falls outside the region of validity of our perturbative treatment ($p - 1 \ll 1$), so the specific values in this plot are not exactly accurate, but the qualitative features are correct. The unstable sidebands satisfy $\delta\omega > \delta\omega_{cr}$, and so their separation will be even greater than for the FM instability. The original RNGH instability is precisely this AM instability, in a traveling-wave laser. For traveling waves, the Lorentzian gain is clamped at threshold regardless of the diffusion parameter, so the instability can only be reached by the parametric enhancement of AM sidebands.

To access both the FM and AM instability regimes experimentally, we need to tune the strength of the PG. The electron diffusivity can be reduced by lowering the temperature, and indeed temperature has a strong effect on the emission spectra of QCLs [102], although the effect is not yet well-understood. In this work, we choose to manipulate the PG by adjusting the facet reflectivities. Increasing the disparity of the reflectivities of the two mirrors reduces the contrast of the standing-wave, because the wave traveling from the higher to the lower-reflectivity facet becomes larger than the counter-propagating wave [123]. For a sufficiently large disparity, the incoherent gain contribution is small enough that the laser can only undergo the AM instability. In a practical sense, engineering the facet coatings allows one to transform a standing-wave cavity into more of a traveling-wave cavity. It is for this reason that we chose to study an HR/AR coated laser, where the AR coating has as low a reflectivity as
current technology allows, to maximize the cavity asymmetry. The full mathematical treatment of mirrors with non-unity reflectivity is complicated by the fact that the spatial modes $\gamma_m(z)$ are no longer orthogonal, and also that the $\gamma_m(z)$ and the longitudinal overlap factors $\Gamma$ vary with the pumping. This theory will be presented in future work.

5.5 Discussion

In order for a mode to oscillate, it must satisfy two conditions: 1) the roundtrip gain must equal the loss, and 2) the roundtrip phase must equal a multiple of $2\pi$. Our theory in Sec. 5.4 has treated only the gain condition. The same approach was taken in the description of the original RNGH instability [79–83]; the underlying assumption is that the cavity modes are densely spaced, so that a pair of sidebands that satisfies the instability condition for the gain will always be “close enough” to two cavity modes that satisfy the phase condition. However, the experimental and theoretical developments of the last decade concerning optical parametric oscillation in externally pumped microresonators have shown that the phase condition has a large effect on the oscillation threshold and sideband spacing [121]. In microresonator experiments, the detuning between the external pump frequency and the center frequency of the cold cavity mode is a degree of freedom that must be precisely controlled to achieve the lowest possible instability threshold. In a laser this detuning is not a degree of freedom, but it can vary with the pumping and should be properly accounted for. A parameter that has no analogy in microresonators is $\Delta$, the detuning between the lasing mode $\omega_0$ and the two-level resonance $\omega_{ba}$, which also varies with the pumping
and is difficult to control in experiments. To precisely predict the instability threshold would require knowledge of both of these detunings, as well as the GVD.

At this stage, the simplest and most important application of the theory is to help determine whether the observed sidebands are parametrically enhanced or suppressed. Because the theory assumes end mirrors with unity reflectivity, we can only expect Eq. 5.19 to apply reasonably well to the uncoated QCLs. For each device, $\gamma_D$ is calculated using the theoretical value of $T_{up}$ (calculated from the bandstructure) and the diffusion constant $D = 77 \text{ cm}^2/\text{s}$ [116], giving $\gamma_D = 0.4$ (DS-3.8), 0.49 (TL-4.6), and 0.93 (LL-9.8). For these large values of $\gamma_D$ the PG is strong, and we find from numerically solving Eq. 5.19 that the FM instability has a lower threshold than the AM instability, regardless of the value of $T_1$. In Sec. 5.9, we show that the theory predicts sideband spacings $\delta\omega_{sb}$ that are consistent with the experimental observations, but underestimates the instability threshold $p_{sb}$. We attribute this discrepancy to the aforementioned detunings and GVD that our theory neglects.

A more direct method to discriminate between the parametric suppression and enhancement regimes is to compare the observed sideband spacing $\delta\omega_{sb}$ to the crossing frequency $\delta\omega_{cr}$. If $\delta\omega_{sb} < \delta\omega_{cr}$, the sidebands are parametrically suppressed and therefore the FM instability has the lower threshold. Thus, we reason that the sidebands must be FM because the AM state would be an unstable equilibrium; an AM waveform, if slightly perturbed, would evolve to an FM waveform. Similarly, if $\delta\omega_{sb} > \delta\omega_{cr}$, the sidebands are parametrically enhanced, so FM sidebands would be an unstable equilibrium and we conclude that they must be AM. Notably, this reasoning depends only on the behavior of $\chi^{(3)}$ as a function of $\delta\omega$; it is therefore
independent of GVD and can be applied to both the uncoated and HR/AR lasers. To calculate $\delta \omega_{cr}$ from Eq. 5.17, we use our measured values of $T_2$ but still need an estimate for the gain recovery time $T_1$. Pump-probe experiments [124,125] and theory [126] have shown that $T_1$ is around 2 ps. From Eq. 5.17, we see that $\delta \omega_{cr}$ decreases with increasing $T_1$, so if we take $T_1 = 3$ ps as a generous upper bound on the gain recovery time, we establish a lower bound of $\delta \omega_{cr}$ at 2270 GHz (DS-3.8), 1730 GHz (TL-4.6), and 1660 GHz (LL-9.8). The measured values of $\delta \omega_{sb}$ for each uncoated laser–977 GHz (DS-3.8), 1259 GHz (TL-4.6), and 642 GHz (LL-9.8)–are all substantially smaller than the lower bound on $\delta \omega_{cr}$. This is consistent with the prediction that the uncoated lasers have a lower FM instability threshold than AM threshold, and with these two results we are reasonably confident that the uncoated lasers emit parametrically suppressed FM sidebands. In stark contrast, TL-4.6:HR/AR exhibits a large sideband separation of $\delta \omega_{sb} = 2216$ GHz. If we use the accepted value of $T_1$ equal to 2 ps, we find $\delta \omega_{cr} = 2120$ GHz. The observed sideband spacing is slightly larger than $\delta \omega_{cr}$, suggesting the enhancement regime. While a smaller gain recovery time or non-perturbative calculation would raise $\delta \omega_{cr}$ slightly, this is our first hint that TL-4.6:HR/AR emits parametrically enhanced AM sidebands.

The difference in the range of intracavity power over which the harmonic state persists in the uncoated vs. coated lasers, as shown in Fig. 5.2, is additional evidence that the uncoated devices operate in the suppression regime and the HR/AR device operates in the enhancement regime. Here we propose a qualitative explanation of this feature. Consider a laser operating in the suppression regime. While the FM state is more stable than the AM state in this regime because it minimizes the
amount of gain suppression, it would be even more favorable for the laser to emit a
temporally incoherent state, in a sense “washing out” the PPs (in much the same way
that a spatially incoherent state washes out the PG), thus avoiding any suppression
of the gain altogether. We surmise that the observed dense state is precisely such
a temporally incoherent state; by lasing on an incoherent superposition of adjacent
cavity modes, the laser maximizes the amount of incoherent gain that it extracts
while avoiding the large parametric suppression that would afflict a coherent state
with such a small one-FSR spacing. From the measurement of negative differential
resistance in DS-3.8, we know that the dense state indeed extracts more gain than
the harmonic state. This could explain why the uncoated lasers only exhibit the
harmonic state over a small range of current: the laser soon finds a way to transition
from the parametrically suppressed FM state to the favored dense state. The fact
that TL-4.6:HR/AR exhibits the harmonic state over a large current range suggests
that the harmonic state is more stable than the incoherent state, which can only be
true in the parametric enhancement regime. At a sufficiently high current, when the
spectral span of the harmonic state approaches the gain bandwidth, the incoherent
dense state finally becomes favored for its ability to lase on adjacent modes, despite
no longer benefitting from the parametric enhancement.

We have argued that the dense state is a temporally incoherent state that man-
age to avoid parametric gain suppression. If this is the case, why do the uncoated
QCLs choose to emit a harmonic state at all, and not simply jump from the single-
mode state to the dense state as the current is increased? In fact, the spectral hystere-
sis shown in Fig. 5.3(d-f) proves that the dense state is the favored lasing state down
to barely above threshold. However, this state can only be reached by decreasing the current after the laser has already entered the dense state at high current. When the laser starts in a single-mode state and the current is increased, there is clearly a barrier that prevents the transition to the dense state. In general, introducing noise allows a system to overcome energy barriers and explore a larger volume of its state space. It is likely that delayed optical feedback serves as such a noise source, and explains why it is difficult to observe the harmonic state when optical feedback is not eliminated.

5.6 Conclusion

We have experimentally identified the single-mode instability of QCLs, which is characterized by the appearance of sidebands at FP modes not adjacent to the primary lasing mode. We have seen the behavior in QCLs at three different wavelengths, each based on a different active region design, and with both positive and negative GVD. Therefore, the phenomenon is a general feature of the electron-light dynamics of QCLs. The instability is reached due to the combined contributions of an incoherent gain due to the spatial population grating, and a coherent parametric gain due to the temporal population pulsations. Our theory predicts both an FM instability in situations where the incoherent gain contribution is large, and an AM instability when the incoherent gain contribution is small. To explore the second possibility, we coated the QCL facets with an HR and an AR coating to reduce the incoherent gain contribution; indeed, this modification substantially increases the sideband spacing, and it is likely that the waveform is AM. Following the first appearance of
sidebands at the instability threshold, our measurements show that increasing the pumping generates more sidebands which preserve the initial spacing. This suggests that a cw QCL can self-start into a phase-locked harmonic frequency comb, and must be investigated further. We have also placed our observations and theory within historical context, explaining the relation to optically pumped microresonators and the single-mode instability in traveling-wave lasers.

The future direction of this work is clear. At first, we can take guidance from the well-established understanding of microresonators and exploit their similarity with QCLs to further our understanding. The calculation of the instability threshold will be extended to account for GVD, so that the cold-cavity modes are not necessarily equidistant. We must also better understand the nature of the single-mode solution; specifically, how does its detuning from the resonant frequency $\omega_{\text{res}}$, and also its detuning from the cold-cavity mode that it occupies, affect the nature of the instability threshold? We must account quantitatively for the non-unity facet reflections and the precise shape of the mode profile within the cavity. Experimentally, second-order autocorrelation experiments are needed to establish the temporal nature of these short-period waveforms.

5.7 Supplement: Calculating intracavity intensity $|\tilde{E}_0|^2$

from measured output power

In the single-mode regime, the intracavity intensity of the single-mode determines the strength of the parametric interaction with the sideband fluctuations.
Therefore, we would like to calculate the intracavity intensity from the measured output power. We remain true to the distributed loss approximation, for which the output power is given by

\[ P_{\text{out}} = \frac{\alpha_m \langle E^2 \rangle Lwh}{\sqrt{\mu/\epsilon}} \]  

(5.22)

where \( \alpha_m = \ln[1/(R_1 R_2)]/(2L) \), the length, width, and height of the cavity are \( L \), \( w \), and \( h \), and the time-averaged intensity of the single-mode is \( \langle E^2 \rangle = 2|\mathcal{E}_0|^2 \). We are assuming a uniform field intensity in the transverse dimensions, and therefore not worrying about the transverse overlap factor. We can rearrange this equation for the intracavity intensity

\[ |\mathcal{E}_0|^2 \equiv \kappa^2 T_1 T_2 |\mathcal{E}_0|^2 = \frac{2d^2 T_1 T_2 \sqrt{\mu_0/\epsilon_0}}{\hbar^2 \alpha_m Lwh} P_{\text{out}}. \]  

(5.23)

With this equation, we can convert the measured total output power of each laser into the intracavity intensity, using our measured values of the refractive index \( n_{\text{eff}} \) and the dephasing time \( T_2 \), our best estimates for \( d \) and \( T_1 \), and in the case of the HR/AR laser we have used \( R_1 = 1 \), \( R_2 = 0.01 \). The result is plotted in Fig. 5.2(b) as a function of \( J/J_{\text{th}} \). Each curve is color-coded to indicate the range over which the laser operates in a single-mode, harmonic state, or dense state. Note that the quantity \( |\mathcal{E}_0|^2 \) is only meaningful in the single-mode regime, because we are interested in the intensity of the single-mode before the harmonic regime sets in.

The theoretical formula for the intracavity intensity is

\[ |\mathcal{E}_0|^2 = \frac{p - 1}{1 + \gamma_D/2}, \]  

(5.24)

where \( p \equiv w_{eq}/w_{th} \) is the pump parameter. We emphasize that \( p \) is not the same as \( J/J_{\text{th}} \). The slope of \( |\mathcal{E}_0|^2 \) vs. \( p \) is always between 2/3 and 1, depending on the
diffusion parameter $\gamma_D$. The reference line in Fig. 5.2(b) is drawn with a slope of one to indicate that each of the $|\tilde{E}_0|^2$ vs. $J/J_{th}$ curves has a slope greater than one. Therefore, we conclude that $J/J_{th}$ must underestimate $p$. One factor that contributes to this underestimation is the transparency current $J_{\text{trans}}$: a fixed amount of current that must be delivered to the active region simply to raise the inversion from a negative number to zero. To understand this simply, suppose that the equilibrium inversion scales like $w_{eq} \propto J - J_{\text{trans}}$, and that $J_{\text{trans}}$ remains a constant number at threshold and above. Then the pump parameter $p \equiv w_{eq}/w_{th}$ is expressed in terms of $J$ as

$$p = \frac{J - J_{\text{trans}}}{J_{th} - J_{\text{trans}}}$$

(5.25)

For example, suppose that for a laser with $J_{th} = 500$ mA the harmonic state kicks in at 550 mA, or $J/J_{th} = 1.1$. If the transparency current was $J_{\text{trans}} = 250$ mA, (in other words, half of the threshold current, which is reasonable for QCLs), then the pump parameter at the harmonic state onset would be $p = (550 - 250)/(500 - 250) = 1.2$. Thus, $J/J_{th}$ underestimates $p$.

A more rigorous study is required to determine $J_{\text{trans}}$ for each laser, which can be done by measuring many lasers of the same active region but different lengths. Once $J_{\text{trans}}$ is known, the slope of $|\tilde{E}_0|^2$ vs. $p$ should fall between $2/3$ and 1 and in principle a value for $\gamma_D$ can be extracted, allowing one to quantify the amount of diffusion present.
5.8 Supplement: Gain recovery time vs. Upper state lifetime

In a QCL, the upper state lifetime $T_{up}$ tells us how long an electron sits in the upper state before making a nonradiative transition to the lower state. From here, it takes some additional time to travel through the injector region and tunnel into the upper level of the next stage. This additional amount of time is the bottleneck that determines the gain recovery time. The Maxwell-Bloch equations, by making the two-level approximation, cannot account for the full complexity of the QCL, and only provide us with one carrier relaxation time, which we have called $T_1$. This begs the question: does $T_1$ represent the upper state lifetime or the gain recovery time? The answer is that it depends on what you want to calculate. In the steady-state single-mode regime, we find that the output power and population inversion are functions of $T_1$ due to diffusion; here, we argue that $T_1$ should represent the upper state lifetime $T_{up}$, because $T_{up}$ tells us how much time an electron in the upper state has to diffuse before transitioning to the lower state. It is for this reason that $T_{up}$ appears in the definition of $\gamma_D$, $\gamma_D = (1 + 4k^2DT_{up})^{-1}$, rather than $T_1$. In dynamical situations, on the other hand, the intensity of the field varies with time and we are interested in the how the population inversion responds. We argue that this response is determined by the gain recovery time, not the upper state lifetime, because the response definitely depends on how long it takes an electron to get from one active stage to the next. To summarize it concisely, the upper state lifetime is used for the calculation of the population grating (PG), but the gain recovery time is used for the calculation of the
population pulsations (PPs). Since the bulk of our manuscript deals with PPs, we chose simply to call $T_1$ the gain recovery time, rather than name a new time scale such as $T_{gr}$, for instance. We hope that this does not confuse the reader.

5.9 Supplement: Numerically calculating the instability threshold

In this section, we demonstrate the predictions of the theory for the three uncoated lasers, and compare the results with the measurements.

Because the theory assumes end mirrors with unity reflectivity, we can only expect Eq. 5.19 to apply reasonably well to the uncoated QCLs. For each device, $\gamma_D$ is calculated using the theoretical value of $T_{up}$ (calculated from the bandstructure) and the diffusion constant $D = 77$ cm$^2$/s [116], giving $\gamma_D = 0.4$ (DS-3.8), 0.49 (TL-4.6), and 0.93 (LL-9.8). For these large values of $\gamma_D$, the incoherent gain increases rapidly with the pumping, and we find from Eq. 5.19 that the FM instability will have a lower threshold than the AM instability, regardless of the value of $T_1$. The gain recovery time $T_1$ of each QCL is not as easily calculable as $T_{up}$ because it depends on a few other time constants of the active region, such as the escape time of the electron from one injector region to the next active region. Therefore, we treat $T_1$ as a variable and calculate the instability threshold $p_{sb}$ and sideband spacing $\delta \omega_{sb}$ as a function of $T_1$. The resulting curves are shown in Fig. 5.8. By comparing the curves with the measured values of $\delta \omega_{sb}$, we can deduce the values $T_1 = 1.83$ ps (DS-3.8), 1.15 ps (TL-4.6), and 0.91 ps (LL-9.8). For these values of $T_1$, the theory predicts an
Figure 5.8: Numerical solutions of the instability threshold obtained by setting the gain $\tilde{g}$ in Eq. 5.19 equal to the loss $\tilde{\ell}$, yielding both (a) the sideband separation $\delta\omega_{sb}T_2$ and (b) the pumping $p_{sb}$. The experimentally measured values of $\delta\omega_{sb}T_2$ are compared to the theory to infer $T_1/T_2$, which also gives the theoretical prediction for the instability threshold $p_{sb}$.
Chapter 5: The quantum cascade laser as a self-pumped parametric oscillator

instability threshold of $p_{sb} = 1.02$ (DS-3.8), 1.09 (TL-4.6), and 1.04 (LL-9.8). It is encouraging that these fitted values of $T_1$ are close to the accepted value of the QCL gain recovery time, which has been shown by pump-probe experiments [124,125] and theory [126] to be around 2 ps. However, the predicted $p_{sb}$ is significantly lower than the measured values $J_{sb}/J_{th} = 1.12$ (DS-3.8), 1.17 (TL-4.6), and 1.14 (LL-9.8), and the discrepancy is made worse by the fact that $J/J_{th}$ is likely an underestimate of $p$ (see the discussion in Sec. 5.7). The fact that the theory underestimates the instability threshold is perhaps not surprising, as we have only made sure that one of the two necessary conditions for sideband oscillation is satisfied (gain, not phase). We hope that future work which accounts for the detuning $\Delta$, the detuning between the lasing mode and the cold cavity mode it occupies, and GVD can accurately predict the instability threshold, which would be a milestone in the understanding of lasers, and also yield a novel laser characterization method of lifetimes and diffusion rates by comparing measured values of $p_{sb}$ and $\delta \omega_{sb}$ to an established theory.
Chapter 6

Theory of parametric gain in lasers

6.1 Introduction

We provide a detailed theoretical derivation of the results that were used in Chapter 5.

6.2 Single-mode solution

This section gives a more detailed derivation of the single-mode solution, including the intracavity power as a function of pumping, and the population inversion as a function of position and pumping.

For a two level system with upper state $|a\rangle$ and lower state $|b\rangle$, the material
equations in the non-rotating frame and the field equation are

\[
\frac{d\rho_{ab}}{dt} = -i\omega_{ba}\rho_{ab} - \frac{id}{\hbar} E(t)w - \frac{\rho_{ab}}{T_2}, \tag{6.1}
\]

\[
\frac{dw}{dt} = \frac{-2id}{\hbar} E(t)(\rho_{ab} - \rho_{ab}^*) + \frac{w_{eq} - w}{T_1} + D\frac{\partial^2 w}{\partial z^2}, \tag{6.2}
\]

\[
\frac{\partial^2 E}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} = Nd\mu \frac{\partial^2}{\partial z^2} (\rho_{ab} + \rho_{ab}^*). \tag{6.3}
\]

We emphasize that these equations are in the non-rotating frame, whereas the Maxwell-Bloch equations we have used in Chapter 5 were already in the rotating frame and the RWA had already been applied. However, since we are here dealing with two counter-propagating waves, we chose to more closely follow the approach in [102]. We make the following ansatzes:

\[
E(z, t) = \frac{1}{\sqrt{2}} \left[ E_R(z, t)e^{-i(\omega t - k z)} + E_L(z, t)e^{-i(\omega t + k z)} + c.c. \right], \tag{6.4}
\]

\[
\rho_{ab}(z, t) = \eta_R^*(z, t)e^{-i(\omega t - k z)} + \eta_L^*(z, t)e^{-i(\omega t + k z)}, \tag{6.5}
\]

\[
w(z, t) = w_{DC}(z, t) + w_2(z, t)e^{i2kz} + w_2^*(z, t)e^{-i2kz}. \tag{6.6}
\]

(We use the subscript “DC” rather than “0” for the spatial average of the population inversion, \(w_{DC}\), because the subscript 0 is used throughout the text to refer to the primary mode. No such ambiguity occurs for the subscript “2.”) Plugging the ansatzes into the differential equations, and making the RWA as well as the slowly-varying
envelope approximation (SVEA) yields the following equations:

\[
\frac{d\eta_R^*}{dt} = -\frac{i\kappa}{2\sqrt{2}}(\mathcal{E}_R w_{DC} + \mathcal{E}_L w_2) - \left( \frac{1}{T_2} + i\Delta \right) \eta_R^* \quad (6.7)
\]

\[
\frac{d\eta_L^*}{dt} = -\frac{i\kappa}{2\sqrt{2}}(\mathcal{E}_L w_{DC} + \mathcal{E}_R w_2^*) - \left( \frac{1}{T_2} + i\Delta \right) \eta_L^* \quad (6.8)
\]

\[
\frac{dw_{DC}}{dt} = \frac{i\kappa}{\sqrt{2}}(\mathcal{E}_R \eta_R + \mathcal{E}_L \eta_L - \text{c.c.}) + \frac{w_{eq} - w_{DC}}{T_1} \quad (6.9)
\]

\[
\frac{dw_2}{dt} = \frac{i\kappa}{\sqrt{2}}(\mathcal{E}_R \eta_R^* - \mathcal{E}_L \eta_L^*) - \frac{w_2}{T_1} - 4k^2 Dw_2 \quad (6.10)
\]

\[
\frac{1}{c} \frac{\partial \mathcal{E}_R}{\partial t} = -\frac{\partial \mathcal{E}_R}{\partial z} + \frac{i\sqrt{2}\alpha}{\kappa T_2} \eta_R^* - \frac{\ell_0}{2} \mathcal{E}_R \quad (6.11)
\]

\[
\frac{1}{c} \frac{\partial \mathcal{E}_L}{\partial t} = +\frac{\partial \mathcal{E}_L}{\partial z} + \frac{i\sqrt{2}\alpha}{\kappa T_2} \eta_L^* - \frac{\ell_0}{2} \mathcal{E}_L \quad (6.12)
\]

where \( \kappa = 2d/h, \alpha = N\omega T_2d^2\sqrt{\mu/\epsilon}/h \) is the Beer absorption coefficient of the material, and \( \Delta = \omega_{ba} - \omega \) is the detuning of the field from the atomic resonance frequency. The loss term \( \ell_0 \) has been added to the field equation heuristically, and in this context it represents only the waveguide loss.

We solve for the single-mode solution by setting the time-derivatives to zero and the slowly-varying envelope functions to be constants. In doing so, we are now making the distributed loss approximation because we are not allowing the fields to grow in space. Thus, \( \ell_0 \) must now be taken to be the total loss, waveguide plus mirror loss. We take \( \Delta = 0 \) for simplicity, because the single-mode will lase very close to the peak of the gain spectrum. We denote the steady-state field amplitudes by \( \mathcal{E}_R = \mathcal{E}_L = \mathcal{E}_0 \) and find the LI curve

\[
|\tilde{\mathcal{E}}_0|^2 = \frac{p - 1}{1 + \gamma_D/2} \quad (6.13)
\]

where \( \gamma_D = (1 + 4k^2 DT_1)^{-1} \) is the diffusion parameter. Based on the discussion in Sec. 5.8, however, we know \( T_1 \) represents the upper state lifetime \( T_{up} \), so we define
\[
\gamma_D = (1 + 4k^2 DT_{up})^{-1}.
\]
The steady-state population \( w_0(z) \) is given by
\[
w_0(z) = w_{th}\left[ 1 + \frac{\gamma_D}{2} \frac{p-1}{1+\gamma_D/2} - \gamma_D \frac{p-1}{1+\gamma_D/2} \cos(2k_0z) \right].
\] (6.14)

### 6.3 Frequency mixing due to population pulsations

This section gives a more detailed derivation of the population pulsations, and demonstrates how to include nonzero detuning \( \Delta \) and GVD into the formalism.

We begin by imagining a small volume of dipoles subject to a spatially uniform \( E \)-field to develop an understanding of the nonlinear effects caused by the Bloch dynamics. The electric field is given by
\[
E(t) = \mathcal{E}(t)e^{i\omega t} + c.c.
\] (6.15)

The Bloch equations in the rotating wave approximation are
\[
\dot{\sigma} = \left( i\Delta - \frac{1}{T_2} \right) \sigma + \frac{i\kappa}{2} w\mathcal{E}
\] (6.16)
\[
\dot{w} = i\kappa(\mathcal{E}^*\sigma - \mathcal{E}\sigma^*) - \frac{w - w_{eq}}{T_1}
\] (6.17)

where \( \sigma \) is the off-diagonal element of the density matrix in the rotating frame, \( w \) is the population inversion (positive when inverted), \( \Delta = \omega_{ba} - \omega \) is the detuning between the applied field and the resonant frequency of the two-level system, \( T_1 \) is the (longitudinal) population relaxation time, \( T_2 \) is the (transverse) dephasing time, \( \kappa \equiv 2d/\hbar \) is the coupling constant where \( d \) is the dipole matrix element (assumed to be real) and \( \hbar \) is Planck’s constant, and \( w_{eq} \) is the equilibrium population inversion in the absence of any electric field which is determined by the pumping. (Note that these equations are identical to Eqs. 3.19(a)-(c) in [117], except that we have allowed
to be complex and left the off-diagonal component of the density matrix in complex notation rather than writing $\sigma = (u+iv)/2$. With these conventions, the macroscopic polarization $P$ (dipole moment per volume) in a region with a volume density of $N$ dipoles is given by

$$P(t) = N d \sigma e^{i\omega t} + c.c.$$  \hspace{1cm} (6.18)

First, we consider the effect of a monochromatic field at frequency $\omega$, obtained from Eqs. 6.16-6.17 by setting $\mathcal{E}(t) = \mathcal{E}_0$ and all time derivatives to zero. The result is a steady-state polarization $\sigma_0$ and population inversion $w_0$ given by

$$\sigma_0 = \frac{i\kappa T_2}{2(1-i\Delta T_2)} w_0 \mathcal{E}_0$$  \hspace{1cm} (6.19)

$$w_0 = \frac{w_{eq}}{1 + \kappa^2 T_1 T_2 |\mathcal{E}_0|^2 / (1 + (\Delta T_2)^2)}$$  \hspace{1cm} (6.20)

Note that the population inversion $w_0$ is saturated as the field strength $\mathcal{E}_0$ increases: this is responsible for saturable loss (when $w_{eq} < 0$) and saturable gain (when $w_{eq} > 0$).

N.B. In our equations so far, we have said the frequency of the field is $\omega$. Later on, we refer to the primary mode frequency as $\omega_0$. For our purposes here, $\omega$ and $\omega_0$ are interchangeable. In future work, this will not be the case. In analogy with the theory developed for microresonators [127], we plan in future work to adopt the convention that $\omega_0$ represents the center-frequency of the cold-cavity mode that is lasing. However, the lasing frequency $\omega$ can be detuned from this cold-cavity resonance due to small frequency-pulling effects when one accounts for the hot cavity. This detuning is an important parameter in microresonators, where the pump frequency $\omega$ can be controllably tuned away from $\omega_0$, allowing one to compensate for GVD and
optimize comb generation. We have not accounted for such a detuning in our work, and therefore $\omega$ and $\omega_0$ are interchangeable.

### 6.3.1 Two-frequency operation

Next, we consider the $E$-field

$$E = E_0 + E_+ e^{i\delta \omega t}$$

which consists of the strong field $E_0$ at frequency $\omega$ superposed with the much weaker field $E_+$ detuned from $\omega$ by $\delta \omega$. A polarization will of course be induced at $\omega + \delta \omega$. However, a polarization at $\omega - \delta \omega$ also results due to the beat note at $\delta \omega$ which modulates the intensity: the resulting modulation of the population inversion with time (i.e., a population pulsation) leads to nonlinear frequency mixing. We express the full polarization as

$$P(t) = \sum_{m=-,0,+} P_m e^{i\omega_m t} + c.c.$$  \hspace{1cm} (6.22)

where $\omega_+ \equiv \omega + \delta \omega$ and $\omega_- \equiv \omega - \delta \omega$. We can solve for the polarization as done in [119], keeping only terms to first order in the weak field $E_+$, which gives

$$P_0 = \frac{i e}{\omega_0} \tilde{\alpha} w_0 \frac{E_0}{1 - i \Delta T_2}$$  \hspace{1cm} (6.23)

$$P_+ = \frac{i e}{\omega_0} \tilde{\alpha} w_0 \left[ \frac{E_+}{1 - i(\Delta - \delta \omega) T_2} + \Lambda_+^* \tilde{E}_0^* E_+ \right]$$  \hspace{1cm} (6.24)

$$P_- = \frac{i e}{\omega_0} \tilde{\alpha} w_0 \Lambda_- \tilde{E}_0^* E_+$$  \hspace{1cm} (6.25)
where

\[ \Lambda_+^+ = \frac{-(1 + i\delta \omega T_2/2)(1 + i(\Delta + \delta \omega)T_2)[1 - i(\Delta - \delta \omega)T_2]}{(1 + i\Delta T_2)(1 + i\delta \omega T_1)[1 + i(\Delta + \delta \omega)T_2][1 - i(\Delta - \delta \omega)T_2] + (1 + i\delta \omega T_2)|\tilde{E}_0|^2} \]

(6.26)

\[ \Lambda_-^+ = \frac{-(1 - i\delta \omega T_2/2)}{(1 - i\Delta T_2)(1 - i\delta \omega T_1)[1 + i(\Delta + \delta \omega)T_2][1 - i(\Delta - \delta \omega)T_2] + (1 - i\delta \omega T_2)|\tilde{E}_0|^2} \]

(6.27)

are the self-mixing and cross-mixing coupling coefficients, respectively. We consider the dipoles to be embedded in a host medium of permittivity \( \epsilon \) and permeability \( \mu \). (We adopt the convention of \cite{45}: \( \epsilon, \mu \) and the speed of light \( c = 1/\sqrt{\epsilon \mu} \) always take their values in the background host medium.) Many of the material properties of the two-level system are lumped into the “Beer loss rate”

\[ \bar{\alpha} = \frac{Nd^2T_2\omega_0c\sqrt{\mu/\epsilon}}{\hbar}, \]

(6.28)

which is related to the more familiar Beer absorption coefficient \( \alpha \) (with units of inverse length) that appears in Beer’s law of absorption by \( \bar{\alpha} = \alpha c \). (Note, however, that in our expressions for the polarization due to the two-level system, all factors of \( \epsilon \) and \( \mu \) drop out; that is, these expressions do not contain the polarization contributions due to the background medium.) The central mode amplitude \( \mathcal{E}_0 \) has been normalized such that \( \tilde{\mathcal{E}}_0 \equiv \kappa \sqrt{T_1T_2}\mathcal{E}_0 \). Note that \( \mathcal{P}_0 \) is unaffected to first order in \( \mathcal{E}_+ \). The polarization \( \mathcal{P}_+ \) comes from two contributions. First, there is the linear contribution from the Lorentz oscillator which \( \mathcal{E}_+ \) would induce even in the absence of the strong field \( \mathcal{E}_0 \). Second, there is a contribution due to the PP which is described by the term \( \Lambda_+^+ \). The term \( \mathcal{P}_- \) is due solely to the PP and is governed by \( \Lambda_-^+ \). Note that the full
polarization is directly proportional to the steady-state population inversion \( w_0 \); this will be important when we generalize our results to standing-wave cavities, where \( w_0 \) varies with position.

Now that we have the polarization, we can calculate the gain seen by the sideband field. We define the gain \( \bar{g} \) (with dimension of frequency) of the sideband as the power density generated at \( \omega + \delta \omega \) by the interaction of the field with the dipoles—considering only field and polarization terms oscillating at \( \omega + \delta \omega \)—divided by the energy density of the exciting sideband field, or

\[
\bar{g}_+ = \frac{(E \dot{P})_+}{2\epsilon |\mathcal{E}_+|^2} = \frac{i\omega_+ (\mathcal{E}_+ \mathcal{P}^*_+ - \mathcal{E}^*_+ \mathcal{P}_+)}{2\epsilon |\mathcal{E}_+|^2}.
\]

Let us consider the case of zero detuning, \( \Delta = 0 \), which simplifies the mathematical expressions considerably and is a prerequisite to understanding the case of non-zero detuning. Under this simplified scenario, we denote the self-mixing coefficient \( \Lambda^+ \) by \( \Lambda \), where

\[
\Lambda = \frac{-(1 + i\delta \omega T_2/2)}{(1 + i\delta \omega T_1)(1 + i\delta \omega T_2)^2 + (1 + i\delta \omega T_2)|\mathcal{E}_0|^2},
\]

and it is simple to show that the cross-coupling coefficient \( \Lambda^- \) is simply \( \Lambda^* \). The gain of the sideband field is found to be

\[
\bar{g}_+ = \bar{\alpha} w_0 \left[ \frac{1}{1 + (\delta \omega T_2)^2} + \text{Real}(\Lambda)|\mathcal{E}_0|^2 \right].
\]

(We have used \( (\omega + \delta \omega)/\omega_{ba} \approx 1 \).) Thus, the gain can be nicely divided up into a contribution from the Lorentz oscillator and a contribution from the PP. All of this is proportional to \( \bar{\alpha} w_0 \); \( \bar{\alpha} \) gives you the gain of a weak field tuned to line-center in
a perfectly inverted medium (or alternatively, the loss seen by a weak field tuned to line-center in a material in its ground state), and \( w_0 \) gives you the expectation value of finding an electron in the excited state (equal to 1 when excited, -1 when in the ground state, and 0 at transparency). Note that \( \text{Real}(\Lambda) \) can be positive or negative, which we will discuss shortly.

### 6.3.2 Three-frequency operation

Of course, the polarization created at \( \omega - \delta \omega \) will create a field at that frequency, which is precisely why in the experiments we always observe the two sidebands appearing simultaneously. One sideband cannot exist in isolation when the mixing terms naturally couple them together. Therefore, we need to consider the field

\[
\mathcal{E} = \mathcal{E}_0 + \mathcal{E}_+ e^{i\delta \omega t} + \mathcal{E}_- e^{-i\delta \omega t}.
\]

The polarization at each sideband frequency now contains a Lorentzian term, a self-mixing term, and a cross-mixing term:

\[
P_+ = \frac{i \epsilon}{\omega_{ba}} \tilde{\alpha} w_0 \left[ \frac{\mathcal{E}_+}{1 - i(\Delta - \delta \omega)T_2} + \Lambda_{+}^{\pm} \tilde{\mathcal{E}}_0 \mathcal{E}_+^* + \Lambda_{-}^{\pm} \tilde{\mathcal{E}}_0 \mathcal{E}_-^* \right]
\]

\[
P_- = \frac{i \epsilon}{\omega_{ba}} \tilde{\alpha} w_0 \left[ \frac{\mathcal{E}_-}{1 - i(\Delta + \delta \omega)T_2} + \Lambda_{-}^{\pm} \tilde{\mathcal{E}}_0 \mathcal{E}_-^* + \Lambda_{+}^{\pm} \tilde{\mathcal{E}}_0 \mathcal{E}_+^* \right]
\]

where \( \Lambda_-^{\pm} \) and \( \Lambda_+^{\pm} \) are obtained by making the substitution \( \delta \omega \rightarrow -\delta \omega \) in the expressions for \( \Lambda_+^{\pm} \) and \( \Lambda_-^{\pm} \), respectively, given in Eqs. 6.26-6.27.

Let us again focus on the case \( \Delta = 0 \), for which the polarization at each sideband
simplifies to
\[ P_+ = \frac{i\epsilon}{\omega_{ba}} \tilde{\alpha} w_0 \left[ \frac{\mathcal{E}_+}{1 + i\delta \omega T_2} + \Lambda \tilde{E}_0 \tilde{E}_0^* \mathcal{E}_+ + \Lambda \tilde{E}_0 \tilde{E}_0^* \mathcal{E}_- \right] \]  \quad (6.36)
\[ P_- = \frac{i\epsilon}{\omega_{ba}} \tilde{\alpha} w_0 \left[ \frac{\mathcal{E}_-}{1 - i\delta \omega T_2} + \Lambda^* \tilde{E}_0 \tilde{E}_0^* \mathcal{E}_- + \Lambda^* \tilde{E}_0 \tilde{E}_0^* \mathcal{E}_+ \right] , \]  \quad (6.37)

where \( \Lambda \) is simply \( \Lambda_+^\pm \) evaluated for \( \Delta = 0 \). We see the nice property that when \( \Delta = 0 \), \( \Lambda_+^+ = \Lambda_+^- (\equiv \Lambda) \), and \( \Lambda_-^- = \Lambda_+^+ (\equiv \Lambda^*) \); in other words, the self- and cross-mixing coupling coefficients are equal.

The gain \( \tilde{g}_+ \) of the positive sideband is
\[ \tilde{g}_+ = \tilde{\alpha} w_0 \left\{ \frac{1}{1 + (\delta \omega T_2)^2} + \text{Real} \left[ \Lambda |\tilde{E}_0|^2 \left( 1 + \frac{\tilde{E}_0^2 \mathcal{E}_-^*}{|\tilde{E}_0|^2 \mathcal{E}_+} \right) \right] \right\} , \]  \quad (6.38)
and a similar expression holds for the minus sideband. This equation tells us that the PP contribution to the gain depends on the phase and amplitude relationships of \( \mathcal{E}_0, \mathcal{E}_-, \) and \( \mathcal{E}_+ \), which is not too surprising because the amplitude of the PP itself is sensitive to these parameters. Without loss of generality, we can take \( \mathcal{E}_0 \) to be real. If \( \mathcal{E}_+ = \mathcal{E}_+^* \), then the two sidebands’ contributions to the beat note at \( \delta \omega \) add constructively, resulting in a field whose amplitude modulation (AM) is twice the strength of a field with only one sideband. If \( \mathcal{E}_+ = -\mathcal{E}_+^* \), then the two sidebands’ contributions to the beat note at \( \delta \omega \) destructively cancel and there is no longer any amplitude modulation at frequency \( \delta \omega \). We refer to such a field as frequency-modulated (FM). We see from Eq. 6.38 that the AM sidebands therefore experience a PP contribution to the gain that is twice as large as the single sideband case, while the FM sidebands experiences only the background Lorentzian gain, consistent with the fact that there is no PP in this case. We summarize this with the formula for the gain \( \tilde{g} \) of each
Chapter 6: Theory of parametric gain in lasers

sideband for the case of equal-amplitude sidebands ($|\mathcal{E}_+| = |\mathcal{E}_-|$),

$$
\bar{g} = \bar{\alpha} w_0 \left[ \frac{1}{1 + (\delta \omega T_2)^2} + \text{Real}(\lambda)|\mathcal{E}_0|^2 \begin{cases} 
2 & \text{AM}; \\
0 & \text{FM};
\end{cases} \right].
$$

(6.39)

Note that for a superposition of AM and FM, the gain due to the PP will fall between 0 and 2 times the factor $\text{Real}(\lambda)|\mathcal{E}_0|^2$.

6.4 Instability threshold in lasers

This section gives a more detailed derivation of the instability threshold, and demonstrates that the gain seen by the sidebands is due to a contribution from the population grating and another from the population pulsations.

When a continuous-wave (cw) laser is pumped at its lasing threshold, only a single frequency of light—the one nearest the gain peak that also satisfies the roundtrip phase condition—has sufficient gain to overcome the roundtrip loss and begins to lase. As the pumping is increased, the single-mode solution yields to multimode operation; this is known as the single-mode instability. Our goal is to determine 1) how hard to pump the laser to reach the single-mode instability and 2) which new frequencies start lasing.

Consider a laser pumped above threshold that is lasing on a single-mode, which we refer to as the primary or central mode. If another mode is to lase, it must be seeded by a spontaneously generated photon at a different frequency. This photon will necessarily create a beat note through its coexistence with the primary mode, resulting in a population pulsation. The gain seen by the new frequency must therefore account for this parametric gain in addition to the background Lorentzian gain. Furthermore,
Chapter 6: Theory of parametric gain in lasers

the PP couples the sideband to the symmetrically detuned sideband frequency on
the other side of the primary mode, so we should in general assume the presence of
both sidebands. Because the instability threshold depends on the cavity geometry,
we will consider a traveling-wave laser as well as a standing-wave laser. In both
cases, the strategy is the same. First, we solve for the single-mode intensity $E_0$ and
the population inversion $w_0(z)$ as a function of the pumping, entirely neglecting the
sidebands. Knowing this, we can then calculate the sideband gain in the presence of
the primary mode.

We start with the wave equation

$$\frac{\partial^2 E}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} = \mu \frac{\partial^2 P}{\partial t^2}.$$ (6.40)

Following the approach used to calculate the optical parametric oscillation threshold
in optically pumped microresonators [127], we expand the field in terms of the cold
cavity modes,

$$E(z, t) = \sum_{m=-0,+} E_m(t) \Upsilon_m(z)e^{i\omega_m t} + c.c.$$ (6.41)

The spatial modes obey the normalization condition

$$\frac{1}{L} \int_0^L dz \, |\Upsilon_m(z)|^2 = 1.$$ (6.42)

When group velocity dispersion (GVD) is non-zero, the two modes $\omega_+$ and $\omega_-$ will
not be equidistant from $\omega_0$. We have also assumed that the spatial and temporal
dependence of the modes can be separated. This is a good approximation in the case
of a laser, because we know the intracavity field will be sharply resonant at the modes.

The spatial variation of the polarization can be described by making the substitution
$E_m \rightarrow E_m \Upsilon_m(z)$ and $w_0 \rightarrow w_0(z)$ into the polarization Eqs. 6.34-6.35, which results in
the polarization

$$P(z, t) = \sum_{m=-,0,+} P_m(z, t) e^{i\omega_m t} + c.c. \quad (6.43)$$

where

$$P_+(z, t) = \frac{i\epsilon}{\omega_{ba}} \tilde{w}_0(z) \left[ \frac{\mathcal{E}_+ Y_+^z(z)}{[1 - i(\Delta - \delta\omega)T_2]} + \Lambda_+^+(z)|\gamma_0(z)|^2\gamma_+(z)|\tilde{E}_0|^2\mathcal{E}_+ \right. + \left. \Lambda_+^-(z)|\gamma_0(z)|^2\gamma_-(z)|\tilde{E}_0|^2\mathcal{E}_- \right] \quad (6.44)$$

$$P_-(z, t) = \frac{i\epsilon}{\omega_{ba}} \tilde{w}_0(z) \left[ \frac{\mathcal{E}_- Y_-^z(z)}{[1 - i(\Delta + \delta\omega)T_2]} + \Lambda_-^+(z)|\gamma_0(z)|^2\gamma_-(z)|\tilde{E}_0|^2\mathcal{E}_- \right. + \left. \Lambda_-^+(z)|\gamma_0(z)|^2\gamma_+(z)|\tilde{E}_0|^2\mathcal{E}_+ \right]. \quad (6.45)$$

We have introduced $\tilde{\omega} \equiv 2\omega_0 - \omega_+ - \omega_-$, the deviation of the cold cavity modes from equal spacing. Note that the $\Lambda$s now depend on $z$ due to the term in their denominators dependent on the primary mode amplitude. Because we no longer demand that the two sidebands have the same detuning $\delta\omega$, $\Lambda_+^+$ and $\Lambda_-^+$ should, strictly speaking, be calculated using the detuning $\delta\omega_+ = \omega_+ - \omega_0$, while $\Lambda_-^+$ and $\Lambda_-^-$ should depend on $\delta\omega_+ = \omega_0 - \omega_-$. In practice, we can ignore this difference in the $\Lambda$s; the term $e^{i\tilde{\omega}t}$ captures the most important effect of GVD.

Plugging everything into the wave equation gives

$$\sum_m \left( \frac{d^2 \gamma_m}{dz^2} + \frac{\omega_m^2}{c^2} \gamma_m \right) \mathcal{E}_m e^{i\omega_m t} - \frac{2i}{c^2} \sum_m \omega_m \frac{d\mathcal{E}_m}{dt} \gamma_m e^{i\omega_m t} = \mu \sum_m -\omega_m^2(P_m - P_{m,\text{loss}}) e^{i\omega_m t} \quad (6.46)$$

where the slowly-varying-envelope approximation allowed us to ignore second time derivatives of $\mathcal{E}_m$ on the left-hand side, and first and second derivatives of $\mathcal{E}_m$ on the right-hand side. The spatial modes $\gamma_m(z)$ are chosen so that the first term on the LHS equals zero. The loss of each mode has been added to the equation in the form
of a polarization contribution; we assume each mode has the same linear loss, which can be expressed

\[ P_{\text{m, loss}}(z, t) = \frac{i\epsilon}{\omega_{ba}} \tilde{\ell} \Upsilon_m(z) E_m(t). \] (6.47)

Equation 6.46 couples all of the modes \( E_m \). We can project this equation onto each mode by multiplying by \( \Upsilon_n(z) \) and integrating over the length of the laser cavity, thus taking advantage of the orthonormality of the spatial modes \( \Upsilon_m(z) \), and then equating terms which oscillate at the same frequency (since terms with different frequencies will not affect the time-averaged gain seen by a mode). The result is one equation for the central mode

\[ \hat{E}_0 = \left[ -\frac{\tilde{\ell}}{2} + \frac{\tilde{\alpha}}{2(1 - i\Delta T_2)} \int \frac{dz}{L} w_0(z) |\Upsilon_0(z)|^2 \right] E_0, \] (6.48)

one for the positive sideband

\[ \hat{E}_+ = -\frac{\tilde{\ell}}{2} E_+ + \frac{\tilde{\alpha}}{2} \left[ \frac{E_+}{1 - i(\Delta - \delta\omega)T_2} \int \frac{dz}{L} w_0(z) |\Upsilon_+(z)|^2 \\
+ |\hat{E}_0|^2 E_+ \int \frac{dz}{L} w_0(z) \Lambda^+_+(z) |\Upsilon_0(z)|^2 |\Upsilon_+(z)|^2 \\
+ \tilde{\epsilon}_0^2 E_+ e^{i\delta\omega t} \int \frac{dz}{L} w_0(z) \Lambda^+_+(z) \Upsilon_0(z)^2 \Upsilon^*_+(z) \Upsilon^*_+(z) \right], \] (6.49)

and one for the negative sideband

\[ \hat{E}_- = -\tilde{\ell} E_- + \frac{\tilde{\alpha}}{2} \left[ \frac{E_+}{1 - i(\Delta + \delta\omega)T_2} \int \frac{dz}{L} w_0(z) |\Upsilon_-(z)|^2 \\
+ |\hat{E}_0|^2 E_- \int \frac{dz}{L} w_0(z) \Lambda^-_-(z) |\Upsilon_0(z)|^2 |\Upsilon_-(z)|^2 \\
+ \tilde{\epsilon}_0^2 E_- e^{i\delta\omega t} \int \frac{dz}{L} w_0(z) \Lambda^-_+(z) \Upsilon_0(z)^2 \Upsilon^*_-(z) \Upsilon^*_-(z) \right]. \] (6.50)

These three equations will be used to understand the instability threshold. In general, one must first apply the steady-state condition \( \hat{E}_0 = 0 \) to Eq. 6.48 which, together
with the Bloch equation relating the field to the inversion, will yield the amplitude of the primary mode $E_0$ along with the resulting population inversion $w_0(z)$, both as a function of the pumping $w_{eq}$. (The result will be the same as what we calculated for the single-mode solution in Sec. 6.2.) This information is then used in Eqs. 6.49-6.50 to determine the minimum level of pumping $w_{eq}$ at which a pair of sidebands with detuning $\delta \omega$ experiences more gain than loss. This is the instability threshold.

So far, we have kept Eqs. 6.48-6.50 as general as possible to account for arbitrary spatial profiles, GVD, and detuning $\Delta$ between the lasing mode and the peak of the gain spectrum. From here on we will simplify the problem by taking $\bar{\omega} = 0$ (zero GVD) and $\Delta = 0$, and apply these conditions to the simplest possible traveling-wave and standing-wave cavities.

### 6.4.1 Traveling-wave cavity

For the traveling-wave laser, the spatial modes are

$$\Upsilon_m(z) = e^{-ik_mz} \quad (6.51)$$

so every point in the cavity sees the same intensity. At and above threshold, the population inversion is everywhere saturated to the threshold inversion, so $w_0$ is independent of $z$. For $\Delta = 0$, the inversion is

$$w_0 = w_{th} \equiv \frac{\tilde{\ell}}{\alpha} \quad (6.52)$$

and the intensity of the primary mode is given by

$$|\tilde{\mathcal{E}}_0|^2 = p - 1 \quad (6.53)$$
where we have made use of the normalized primary mode amplitude $\tilde{E}_0 \equiv \kappa \sqrt{T_1 T_2} E_0$, and $p$ is the pumping parameter defined as $p \equiv \omega_{eq}/\omega_{th}$. Because $|\tilde{E}_0|^2$ is independent of $z$, all of the $\Lambda$s are independent of $z$. Furthermore, since both $w_0$ and the $\Lambda$s are independent of $z$, they can be pulled out of the spatial integrals in Eqs. 6.49-6.50.

These integrals are then equal to one, where we have used the zero GVD condition $\varpi = 0$ in order for the cross-overlap integral (the last integral in each equation) to equal one. The sideband equations become

\[
\begin{align*}
\dot{\mathcal{E}}_+ &= -\frac{i}{2} \mathcal{E}_+ + \frac{\bar{\alpha} \omega_{th}}{2} \left[ \frac{\mathcal{E}_+}{1 + i\delta \omega T_2} + \Lambda |\tilde{E}_0|^2 \mathcal{E}_+ + \Lambda |\tilde{E}_0|^2 \mathcal{E}_* \right] \quad (6.54) \\
\dot{\mathcal{E}}_- &= -\frac{i}{2} \mathcal{E}_- + \frac{\bar{\alpha} \omega_{th}}{2} \left[ \frac{\mathcal{E}_-}{1 - i\delta \omega T_2} + \Lambda^* |\tilde{E}_0|^2 \mathcal{E}_- + \Lambda^* |\tilde{E}_0|^2 \mathcal{E}_* \right] \quad (6.55)
\end{align*}
\]

which can be written in matrix form

\[
\begin{pmatrix}
\dot{\mathcal{E}}_+ \\
\dot{\mathcal{E}}_-
\end{pmatrix} =
\begin{pmatrix}
M_+ & R_+ \\
R_- & M_-
\end{pmatrix}
\begin{pmatrix}
\mathcal{E}_+ \\
\mathcal{E}_-
\end{pmatrix}
\]

where

\[
\begin{align*}
M_+ &= M_- = -\frac{i}{2} + \frac{\bar{\alpha} \omega_{th}}{2} \left( \frac{1}{1 + i\delta \omega T_2} + \Lambda |\tilde{E}_0|^2 \right) \quad (6.57) \\
R_+ &= R_- = \frac{\bar{\alpha} \omega_{th}}{2} \Lambda |\tilde{E}_0|^2. \quad (6.58)
\end{align*}
\]

(In the last step, we have finally taken the freedom to choose $\tilde{E}_0$ to be real, which we can do at this point without loss of generality.)

Now, if we assume a solution of the form $\mathcal{E}_\pm \sim e^{\lambda t}$, we find the two solutions for $\lambda$

\[
\lambda = \frac{1}{2} [M_+ + M_- \pm \sqrt{(M_+ - M_-)^2 + 4R_+ R_-}]. \quad (6.59)
\]

The net gain seen by each sideband is given by $\text{Real}(2\lambda)$ (the factor of two is for intensity gain rather than amplitude gain), which includes the gain minus the loss.
Subtracting off the loss, the gain $\bar{g}$ seen by each sideband is

$$\bar{g} = \bar{\alpha} \omega_{th} \left[ \frac{1}{1 + (\delta \omega T_2)^2} + \begin{cases} 2 \text{ Real}(\Lambda)|\tilde{E}_0|^2 & ; \text{AM} \\ 0 & ; \text{FM} \end{cases} \right]$$

(6.60)

where the two solutions correspond to AM and FM sideband configurations. Finally, we recognize that the gain is pinned at threshold, so $\bar{\alpha} \omega_{th} = \bar{\ell}$, and we write down the sideband gain normalized to the loss

$$\frac{\bar{g}}{\bar{\ell}} = \frac{1}{1 + (\delta \omega T_2)^2} + \text{Real}(\Lambda)|\tilde{E}_0|^2 \cdot \begin{cases} 2 & ; \text{AM} \\ 0 & ; \text{FM} \end{cases}$$

(6.61)

When the gain $\bar{g}$ exceeds the loss $\bar{\ell}$, the weak sideband amplitudes experience exponential growth, therefore the single-mode solution becomes unstable. Note that the Lorentzian term is always less than 1. This is a direct result of uniform gain clamping in the traveling-wave laser, which clamps the net gain of the mode at the peak of the Lorentzian to zero, and therefore any mode detuned from the peak will see slightly more loss than gain. FM sidebands therefore never become unstable because they only see the Lorentzian gain. On the other hand, AM sidebands induce a PP and with it a coherent gain term, which can provide enough extra gain on top of the Lorentzian background to allow the sidebands to lase,

$$\frac{\bar{g}_{AM}}{\bar{\ell}} = \frac{1}{1 + (\delta \omega T_2)^2} + 2 \text{ Real}(\Lambda)|\tilde{E}_0|^2.$$  

(6.62)

To get a feel for the sideband gain, we have plotted $\bar{g}_{AM}/\bar{\ell}$ in Fig. 6.1 at various pump strengths $p$ for $T_1/T_2 = 1, 10, \text{ and } 100$. Graphically, we see that at large enough $p$ sidebands will become unstable. Analytically, it is a simple matter to calculate how hard to pump the laser $p$ before the sidebands appear, starting from Eq. 6.62. We
start by replacing $|\tilde{E}_0|^2$ with $p - 1$, and note that this substitution must also be made in $\Lambda$, which implicitly varies with $|\tilde{E}_0|^2$. Then, setting $\bar{g}_{AM}/\bar{\ell}$ equal to one, we can solve a simple quadratic formula for $\delta \omega^2$,

$$(\delta \omega T_2)^2 = \frac{-1 + 3Z(p - 1) \pm \sqrt{[1 - 3Z(p - 1)]^2 - 8Z^2p(p - 1)}}{2Z^2}, \quad (6.63)$$

where $Z \equiv T_1/T_2$. Finally, we must apply some physical reasoning: as $p$ is increased past 1, the sideband gain increases. Right at the moment when the instability threshold is reached, $\delta \omega^2$ must take on a single value. Thus, we set the radical in Eq. 6.63 to zero and solve for $p$. After solving another simple quadratic equation, we find that

$$p = 5 + \frac{3}{Z} \pm 4\sqrt{1 + \frac{3}{2Z} + \frac{1}{2Z^2}}. \quad (6.64)$$

How do we choose between the plus and minus sign? By plugging this expression for $p$ back into Eq. 6.63, it is simple to check that only the plus sign yields real-valued solutions for $\delta \omega$. Thus, we have found the instability threshold, which we denote.
Chapter 6: Theory of parametric gain in lasers

\[ p_{RNGH} = 5 + \frac{3}{Z} + 4\sqrt{1 + \frac{3}{2Z} + \frac{1}{2Z^2}} \]  \hspace{1cm} (6.65)

because it is the well-known instability threshold found by Risken and Nummedal (see Eq. 3.10 in [79]) and Graham and Haken (see Eq. 7.35 in [80]). Plugging this value of \( p \) into Eq. 6.63 yields the value of \( \omega \) of the sidebands when the instability sets in

\[ (\delta \omega_{RNGH} T_2)^2 = \frac{4}{Z^2} + 6 \frac{1}{Z} (1 + \sqrt{1 + \frac{3}{2Z} + \frac{1}{2Z^2}}). \]  \hspace{1cm} (6.66)

One thing to notice is that in the limit \( Z \gg 1 \) (transverse relaxation must faster than longitudinal relaxation), the instability threshold \( p_{RNGH} \to 9 \) from above and \( \delta \omega_{RNGH} T_2 \to \sqrt{12/Z} \).

### 6.4.2 Standing-Wave cavity

As before, we restrict ourselves to the case \( \Delta = 0 \) and \( \bar{\omega} = 0 \). We will see that calculations for the standing-wave cavity are significantly more complicated than for the traveling-wave cavity. The spatial variation of the primary mode causes the inversion \( w_0 \) and the coupling \( \Lambda \) to both depend on \( z \), which makes the integrals more difficult to compute. For this reason, we treat the problem to first order in the primary mode intensity \( |\tilde{E}_0|^2 \), which allows us to compute the integrals analytically. However, the theory can be extended to higher order at will, or the integrals can always be computed numerically.

For the standing-wave laser with perfectly reflecting end mirrors, the spatial profile of each mode is given by

\[ \Upsilon_m(z) = \sqrt{2} \cos(k_m z). \]  \hspace{1cm} (6.67)
Chapter 6: Theory of parametric gain in lasers

The spatial modulation of the intensity is responsible for the spatial modulation of the population inversion $w_0(z)$, though mitigated somewhat by carrier diffusion. We calculated $w_0(z)$ in Sec. 6.2. The result is

$$w_0(z) = w_{th} \left[ 1 + \frac{\gamma_D}{2} \frac{p - 1}{1 + \gamma_D/2} - \frac{\gamma_D}{1 + \gamma_D/2} \cos(2k_0z) \right]. \quad (6.68)$$

to first order in $|\tilde{E}_0|^2$, where $w_{th} = \bar{\ell}/\bar{\alpha}$, and $\gamma_D = (1 + 4k^2DT_{up})^{-1}$ is the diffusion parameter. The spatial variation of the inversion has important consequences. For one, it reduces the power of the laser, which is given by

$$|\tilde{E}_0|^2 = \frac{p - 1}{1 + \gamma_D/2}. \quad (6.69)$$

Secondly, the gain is no longer uniformly clamped by the primary lasing mode, which will allow new modes to lase even in the absence of PPs.

The spatial variation of the primary lasing mode also causes $\Lambda$ to vary with position. In keeping with our approximations, we can expand $\Lambda$ to zeroth order in $|\tilde{E}_0|^2$ because in our equations $\Lambda$ always multiplies $|\tilde{E}_0|^2$, so the final result is first order in $|\tilde{E}_0|^2$. We define the zeroth order expansion of $\Lambda$ to be

$$\chi^{(3)} = \frac{-(1 + i\delta\omega T_2/2)}{(1 + i\delta\omega T_1)(1 + i\delta\omega T_2)^2}, \quad (6.70)$$

where the symbol $\chi^{(3)}$ was chosen to emphasize that this term now plays the role of a third-order nonlinear coefficient.

We start with the sideband Eqs. 6.49-6.50, replace $w_0(z)$ with Eq. 6.14, $\Lambda(z)$ with $\chi^{(3)}$, and keep only terms to first order in $|\tilde{E}_0|^2$. The resulting equation for the
growth of the positive sideband is

\[ \dot{E}_+ = -\frac{\tilde{\ell}}{2} E_+ + \frac{\tilde{\alpha} w_{th}}{2} \left[ 1 + \frac{\gamma_0^2}{2} |E_0|^2 \right] E_+ + \chi^{(3)} |E_0|^2 E_+ \int \frac{dz}{L} \left| \gamma_0(z) \right|^2 \left| \gamma_+(z) \right|^2 \]

and a similar equation can be written down for \( \dot{E}_- \). We define the longitudinal overlap integrals

\[ \Gamma_{\text{self}} = \int_0^L \frac{dz}{L} \left| \gamma_0(z) \right|^2 \left| \gamma_+(z) \right|^2 = 1 \]  

\[ \Gamma_{\text{cross}} = \int_0^L \frac{dz}{L} \gamma_0(z)^2 \gamma_-(z) \gamma_+(z) = 1/2. \]

The implication is that the self-mixing interaction of a sideband with itself, mediated by the primary mode intensity, is twice as large as the cross-mixing interaction of one sideband generating gain for the other sideband, again mediated by the primary mode intensity. This is true only for the cosine-shaped modes that we have assumed, and the overlap integrals will change when the longitudinal spatial profile changes, as when the non-unity reflectivity of the facets is taken into account. The sideband Eqs. 6.49-6.50 become

\[ \dot{E}_+ = -\frac{\tilde{\ell}}{2} E_+ + \frac{\tilde{\alpha} w_{th}}{2} \left[ 1 + \frac{\gamma_0^2}{2} |E_0|^2 \right] E_+ + \Gamma_{\text{self}} \chi^{(3)} |E_0|^2 E_+ + \Gamma_{\text{cross}} \chi^{(3)} |E_0|^2 E_- \]

\[ \dot{E}_- = -\frac{\tilde{\ell}}{2} E_- + \frac{\tilde{\alpha} w_{th}}{2} \left[ 1 + \frac{\gamma_0^2}{2} |E_0|^2 \right] E_- + \Gamma_{\text{self}} \chi^{(3)} |E_0|^2 E_- + \Gamma_{\text{cross}} \chi^{(3)} |E_0|^2 E_+ \]

which we express as

\[
\begin{pmatrix}
\dot{E}_+ \\
\dot{E}_-
\end{pmatrix} = \begin{pmatrix}
M_+ & R_+ \\
R_- & M_-
\end{pmatrix} \begin{pmatrix}
E_+ \\
E_-
\end{pmatrix}
\]

(6.76)
where

\begin{align}
M_+ = M_-^* &= -\frac{\ell}{2} + \frac{\alpha_w \text{th}}{2} \left( \frac{1 + \frac{\gamma_D}{2} |\tilde{E}_0|^2}{1 + i\delta\omega T_2} + \Gamma_{\text{self}} \chi^{(3)} |\tilde{E}_0|^2 \right) \tag{6.77} \\
R_+ = R_-^* &= \frac{\alpha_w \text{th}}{2} (\Gamma_{\text{cross}} \chi^{(3)} |\tilde{E}_0|^2) \tag{6.78}
\end{align}

As we did for the traveling-wave laser, the sideband gain is easily calculated from these two coupled first-order differential equations. Normalizing the gain to the total loss, we find

\[ \frac{\bar{g}}{\ell} = \frac{1 + \frac{\gamma_D}{2} |\tilde{E}_0|^2}{1 + (\delta\omega T_2)^2} + \text{Real} \left[ \chi^{(3)} |\tilde{E}_0|^2 \right] . \left\{ \begin{array}{l}
\Gamma_{\text{self}} + \Gamma_{\text{cross}} = \frac{3}{2} ; \text{ AM} \\
\Gamma_{\text{self}} - \Gamma_{\text{cross}} = \frac{1}{2} ; \text{ FM} 
\end{array} \right. \tag{6.79} \]

There are two things to notice here. As the laser pumping is increased, the term \( \gamma_D |\tilde{E}_0|^2/2 \) grows, and consequently the gain is not clamped at the threshold value. This is due to spatial hole burning, or more precisely, the imperfect overlap of the standing-wave modes together with a finite amount of carrier diffusion. We view this background gain as a Lorentzian-shape whose amplitude increases with the pumping, and is therefore fully capable of pulling the sidebands above threshold, without any additional PP contribution to the gain.

Secondly, the PP contribution to the gain never vanishes. Even when the sidebands are phased such that an FM waveform is emitted from the laser, there is still a PP within the laser cavity. The reason for this is the imperfect overlap of the two sidebands’ spatial modes, which means that at any given position within the cavity, the plus and minus sideband are likely to have different amplitudes. Therefore, even if the two sidebands are phased such that their contributions to the beat note at \( \delta\omega \) destructively interfere with each other, the destruction is not perfect. The amplitude
Chapter 6: Theory of parametric gain in lasers

of the PP varies with position in the cavity, and in locations where the two sideband amplitudes are equal the PP will not exist, but the spatially averaged effect of the FM PP yields the factor of $1/2$ in Eq. 6.79. By the same token, sidebands phased for AM will not fully constructively interfere, yielding a factor of $3/2$ for the PP contribution to the gain rather than the factor $2$, as it would be for the traveling-wave laser.

At first glance, it appears from Eq. 6.79 that the AM sidebands always experience more gain than FM sidebands, because $3/2 > 1/2$. However, Real[$\chi^{(3)}$] can be negative. In particular, Real[$\chi^{(3)}$] is negative for low-frequency sidebands and positive for high-frequency sidebands. Therefore, low-frequency FM sidebands are not as strongly suppressed as low-frequency AM sidebands, while high-frequency AM sidebands are more enhanced than high-frequency FM sidebands. Whether a laser will undergo a single-mode instability that results in FM or AM sidebands depends on the strength of the population grating and the overlap factors, and this was discussed in detail in Chapter 5.
Bibliography


[40] Larry A Coldren, Multi-section tunable laser with differing multi-element mirrors, 1990.


Bibliography


Bibliography


